(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 July 2001 (19.07.2001)

PCT

(10) International Publication Number
WO 01/51622 A2

(51) International Patent Classification[7]: C12N 9/02, 15/82

(21) International Application Number: PCT/EP01/00297

(22) International Filing Date: 11 January 2001 (11.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

00100646.9	13 January 2000 (13.01.2000)	EP
00107001.0	30 March 2000 (30.03.2000)	EP
00109423.4	3 May 2000 (03.05.2000)	EP
00114184.5	13 July 2000 (13.07.2000)	EP
00114912.9	17 July 2000 (17.07.2000)	EP

(71) Applicants *(for all designated States except US)*: **SYNGENTA PARTICIPATIONS AG** [CH/CH]; Schwarzwaldallee 215, CH-4058 Basel (CH). **ROYAL VETERINARY AND AGRICULTURAL UNIVERSITY** [DK/DK]; Thorvaldsensvej 40, DK-1871 Frederiksberg C Copenhagen (DK).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ANDERSEN, Mette, Dahl** [DK/US]; 3833 Nobel Drive Apt. #3422, San Diego, CA 92122 (US). **MÖLLER, Birger, Lindberg** [DK/DK]; Kongstedvej 5, DK-2700 Brönshöj (DK). **NIELSEN, John, Strikart** [DK/DK]; Tobrukvej 6, DK-2770 Kastrup (DK). **WITTSTOCK, Ute** [DE/DE]; Mittelstrasse 18, 07745 Jena (DE). **HANSEN, Carsten, Hörslev** [DK/DK]; Morten Nörholm, Blegdamsvej 106A, 3, DK-2100 Köbenhavn Ø (DK). **HALKIER, Barbara, Ann** [DK/DK]; Nansensgade 43, 1. tv, DK-1366 Copenhagen K. (DK). **MIKKELSEN, Michael, Dalgaard** [DK/DK]; Dambos Vaenge 13, 1. th, DK-2500 Valby (DK).

(74) Agent: **BASTIAN, Werner**; Novartis AG, Patent & Trademark Department Agribusiness, Site Rosental, CH-4002 Basel (CH).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/51622 A2

(54) Title: P450 MONOOXYGENASES OF THE CYP79 FAMILY

(57) Abstract: The invention provides DNA coding for cytochrome P450 monooxygenases of the CYP79 family catalyzing the conversion of an aliphatic or aromatic acid or chain-elongated methionine homologue to the corresponding oxime. Preferred embodiments of the invention are enzymes catalyzing the conversion of L-Valine and L-Isoleucine such as the cassava enzymes CYP79D1 and CYP79D2, enzymes catalyzing the conversion of tyrosine such as the *Triglochin maritima* enzymes CYP79E1 and CYP79E2, enzymes catalyzing the conversion of tryptophan to the corresponding oxime indole-3-acetaldoxine such as the *Arabidopsis thaliana* enzyme CYP79A2 and the *Bassica napus* enzyme CYP79B5, and enzymes catalyzing the conversion of a chain-elongated methionine homologue such as the*Arabidopsis thaliana* enzymes CYP79F1 and CYP79F2. Transgenic expression of said DNA or parts thereof in plants can be used to manipulate the biosynthesis of corresponding glucosinolates or cyanogenic glucosides.

Cell Biol. 12: 1-51, 1993). Cytochrome P450s showing more than 55% identity belong to the same subfamily.

Glucosinolates are amino acid-derived, secondary plant products containing a sulfate and a thioglucose moiety. The occurence of glucosinolates is restricted to the order Capparales and the genus *Drypetes* (Euphorbiales). *C. papaya* is the only known example of a plant containing both glucosinolates and cyanogenic glucosides. The order Capparales includes agriculturally important crops of the *Brassicaceae* family such as oilseed rape and *Brassica* forages and vegetables, and the model plant *Arabidopsis thaliana* L. Upon tissue damage, glucosinolates are rapidly hydrolyzed to biologically active degradation products. Glucosinolates or rather their degradation products defend plants against insect and fungal attack and serve as attractants to insects that are specialized feeders on *Brassicaceae*. The degradation products have toxic as well as protective effects in higher animals and humans. Antinutritional effects such as growth retardation caused by consumption of large amounts of rape seed meal have an economical impact as they restrict the use of this protein-rich animal feed. Anticarcinogenic activity has been documented by pharmacological studies for several degradation products of glucosinolates, e.g. for sulforaphane, a degradation product of 4-methylsulfinylbutylglucosinolate from broccoli sprouts. Metabolic engineering of the biosynthetic pathways of glucosinolates allows to tissue-specifically regulate and optimize the level of individual glucosinolates to improve the nutritional value of a given crop. Besides their occurrence in *A. thaliana*, such glucosinolates are important constituents of *Brassica* crops and vegetables. For example, the major glucosinolate in *B. napus*, the goitrogenic 2-hydroxy-3-butenylglucosinolate, is formed by side-chain modification of 4-methylthiobutylglucosinolate. The occurrence of 2-hydroxy-3-butenylglucosinolate in *B. napus* restricts the use of the protein-rich seed cake as animal feed. Thus availability of biosynthetic genes has great potential for the development of crops with reduced levels of undesirable glucosinolates while retaining glucosinolates with desirable effects, *e.g.* for pest resistance.

To date, more than 100 different glucosinolates have been identified. They are grouped into aliphatic, aromatic, and indolyl glucosinolates, depending on whether they are derived from aliphatic amino acids, phenylalanine and tyrosine, or tryptophan. The amino acid often undergoes a series of chain elongations prior to entering the biosynthetic pathway, and the glucosinolate product is often subject to secondary modifications such as hydroxylations,

P450 CYP79B5 from *Brassica napus*. It shows that CYP79A2 catalyzes the conversion of L-phenylalanine to phenylacetaldoxime, CYP79B2 the conversion of tryptophan to indole-3-acetaldoxime, and CYP79F1 the conversion of chain-elongated methionine homologues such as e.g. homo-, dihomo-, trihomo-, tetrahomo-, pentahomo- and hexahomomethionine to their corresponding aldoximes. It further shows that transgenic *A. thaliana* expressing CYP79A2 or CYP79B2 under control of the CaMV35S promoter accumulate high levels of benzyl- or indoleglucosinolates, respectively, whereas transgenic *Arabidopsis thaliana* expressing CYPF1 can show cosuppression of CYPF1 with a reduced content of glucosinolates derived from chain-elongated methionine homologues and with highly increased levels of chain-elongated methionines such as e.g. dihomo- and trihomomethionine. The data are consistent with the involvement of CYP79A2, CYP79B2 and CYP79F1 in the glucosinolate biosynthesis in *A. thaliana*. The presence of an IAOX producing CYP79 in the biosynthesis of indoleglucosinolates is unexpected since no tryptophan-derived cyanogenic glucosides have been identified and a peroxidase activity has been described in the literature as being involved in indoleglucosinolate biosynthesis. Furthermore, indoleglucosinolates are the products of a recent evolutionary event and are present only in four families in the Capparales order, namely in *Brassicaceae*, *Resedaceae*, *Tovariaceae* and *Capparaceae*. Thus, the possible involvement of IAOX in the biosynthesis of both IAA and indoleglucosinolates would suggest that the nature of the enzyme catalyzing the conversion of tryptophan to IAOX is different from a CYP79 *N*-hydroxylase. The characterization of *CYP79B2 in planta* as well as *in vitro* demonstrates, that oxime production by CYP79 proteins in the biosynthesis of glucosinolates is not restricted to those aromatic amino acids that are also precursors in cyanogenic glucoside biosynthesis. This shows that after diverging away from cyanogenic glucosides, CYP79 proteins developed a new substrate specificity. As a consequence thereof, it is expected that a number of cytochrome P450s of glucosinolate producing plants belonging to the CYP79 family, will turn out to catalyze oxime production from various precursor amino acids in glucosinolate biosynthesis.

Cassava, the most important tropical root crop, contains two cyanogenic glucosides, i.e. linamarin and lotaustralin, in all parts of the plant. Upon tissue disruption said glucosides are degraded with concomitant release of hydrogen cyanide. Acyanogenic cassava plants are not known and attempts to completly eliminate cyanogenic glucosides through breeding have not been successful. Thus, use of cassava products as staple food requires careful

- A DNA coding for a P450 monooxygenase converting an aliphatic or aromatic amino acid or chain-elongated methionine homologue, such as valine, leucine, isoleucine, cyclopentenylglycine, tyrosine, L-phenylalanine, tryptophan, dihomo-, trihomo- or tetrahomomethionine to the corresponding oxime;
- Said DNA coding for a P450 monooxygenase, wherein global alignment of the amino acid sequence of the encoded protein shows at least 40% identity to the amino acid sequence resulting from the global alignment with SEQ ID NO: 1 or SEQ ID NO: 3 or both; SEQ ID NO: 39; or SEQ ID NO: 54 or SEQ ID NO: 70 or both; or at least 50% identity to the amino acid sequence resulting from the global alignment with SEQ ID NO: 9 or SEQ ID NO: 11 or both or SEQ ID NO: 74 or SEQ ID NO: 84 or both.
- Said DNA coding for a P450 monooxygenase having the formula R_1-R_2-R_3, wherein
  - -- R_1, R_2 and R_3 designate component sequences, and
  - -- R_2 consists of 150 to 175 or more amino acid residues the sequence of which is at least 60% identical to an aligned component sequence of SEQ ID NO: 1 or SEQ ID NO: 3; SEQ ID NO: 9 or SEQ ID NO: 11; SEQ ID NO: 54 or SEQ ID NO: 70; SEQ ID NO: 74 or SEQ ID NO: 84; or at least 65% identical to an aligned component sequence of SEQ ID NO: 39.
- A P450 monooxygenase converting an aliphatic or aromatic amino acid or a chain-elongated methionine homologue to the corresponding oxime;
- A method for the isolation of a cDNA coding for a P450 monooxygenase converting an aliphatic or aromatic amino acid or chain-elongated methionine homologue to the corresponding oxime;
- A method for producing purified recombinant P450 monooxygenase converting an aliphatic or aromatic amino acid or chain-elongated methionine homologue to the corresponding oxime; and
- A marker assisted breeding method using at least one oligonucleotide of at least 15 to 20 nucleotides length constituting a component sequence of the DNA according to the present invention, and
- A method for obtaining a transgenic plant comprising stably integrated into its genome DNA comprising at least part of an open reading frame of a P450 monooxygenase converting an aliphatic or aromatic amino acid or chain-elongated methionine homologue to the corresponding oxime. Dependent on the constructs used resulting plants show an altered content or profile of cyanogenic glucosides or glucosinolates.

The enzyme is specific for L-amino acids. It consists of amino acid residues independently selected from the group of the amino acid residues Gly, Ala, Val, Leu, Ile, Phe, Pro, Ser, Thr, Cys, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg and His, and shows at least 40%, preferably 55%, or even more preferably 70% identity to the amino acid sequence resulting from global alignment with either SEQ ID NO: 1 (CYP79D1) or SEQ ID NO: 3 (CYP79D2) or both, which sequences define specific embodiments of the present invention naturally expressed in cassava.

The present invention further discloses a P450 monooxygenase converting an aromatic amino acid such as tyrosine or phenylalanine to the corresponding oxime. The enzyme is specific for L-amino acids. It consists of amino acid residues independently selected from the group of the amino acid residues Gly, Ala, Val, Leu, Ile, Phe, Pro, Ser, Thr, Cys, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg and His, and shows at least 50%, preferably 55%, or even more preferably 70% identity to the amino acid sequence resulting from global alignment with either SEQ ID NO: 9 (CYP79E1) or SEQ ID NO: 11 (CYP79E2) or both, which sequences define specific embodiments of the present invention naturally expressed in *Triglochin maritima*.

The present invention further discloses a P450 monooxygenase converting L-phenylalanine to phenylacetaldoxime. It consists of amino acid residues independently selected from the group of the amino acid residues Gly, Ala, Val, Leu, Ile, Phe, Pro, Ser, Thr, Cys, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg and His, and shows at least 40%, preferably 55%, or even more preferably 70% identity to the amino acid sequence resulting from global alignment with SEQ ID NO: 39 (CYP79A2), which defines a specific embodiment of the present invention naturally expressed in *Arabidopsis thaliana*.

The present invention further discloses a P450 monooxygenase converting tryptophan to indole-3-acetaldoxime. It consists of amino acid residues independently selected from the group of the amino acid residues Gly, Ala, Val, Leu, Ile, Phe, Pro, Ser, Thr, Cys, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg and His, and shows at least 40%, preferably 55%, or even more preferably 70% identity to the amino acid sequence resulting from global alignment with SEQ ID NO: 54 (CYP79B2)) or SEQ ID NO: 70 (CYP79B5), which define specific embodiments of the present invention naturally expressed in *Arabidopsis thaliana* and *Brassica napus*, respectively.

The present invention further discloses a P450 monooxygenase converting an aliphatic amino acid or chain-elongated methionine homologue to the corresponding aldoxime. It consists of amino acid residues independently selected from the group of the amino acid

In general there exist two approaches towards sequence alignment. Dynamic programming algorithms as proposed by Needleman and Wunsch and by Sellers align the entire length of two sequences providing a global alingment of the sequences. The Smith-Waterman algorithm on the other hand yields local alignments. A local alignment aligns the pair of regions within the sequences that are most similiar given the choice of scoring matrix and gap penalties. This allows a database search to focus on the most highly conserved regions of the sequences. It also allows similiar domains within sequences to be identified. To speed up alignments using the Smith-Waterman algorithm programs such as BLAST (Basic Local Alignment Search Tool) and FASTA place additional restrictions on the alignments.

Within the context of the present invention global sequence alignments are conveniently performed using the program PILEUP available from the Genetic Computer Group, Madison, WI.
Local alignments are performed conveniently using BLAST, a set of similarity search programs designed to explore all of the available sequence databases regardless of whether the query is protein or DNA. Version BLAST 2.0 (Gapped BLAST) of this search tool has been made publicly available on the internet (currently http://www.ncbi.nlm.nih.gov/BLAST/). It uses a heuristic algorithm which seeks local as opposed to global alignments and is therefore able to detect relationships among sequences which share only isolated regions. The scores assigned in a BLAST search have a well-defined statistical interpretation. Particularly useful within the scope of the present invention are the blastp program allowing for the introduction of gaps in the local sequence alignments and the PSI-BLAST program, both programs comparing an amino acid query sequence against a protein sequence database, as well as a blastp variant program allowing local alignment of two sequences only. Said programs are preferably run with optional parameters set to the default values.

Additionally, sequence alignments using BLAST can take into account whether the substitution of one amino acid for another is likely to conserve the physical and chemical properties necessary to maintain the structure and function of a protein or is more likely to disrupt essential structural and functional features. Such sequence similarity is quantified in terms of a percentage of 'positive' amino acids, as compared to the percentage of identical amino acids and can help assigning a protein to the correct protein family in border-line cases.

the source for the cytochrome P450 enzyme according to standard proceedures (Sibbesen et al, J. Biol. Chem. 270: 3506-3511, 1995).

Alternatively bacteria like *Escherichia coli* can be used for the recombinant expression of cytochrome P450 enzymes belonging to the CYP79 family. The resulting proteins are unglycosylated. Depending on the particular enzyme studied vector constructs with inserts encoding native or various truncated, extended or modified amino terminal sequences are preferred (Halkier et al, Arch. Biochem. Biophys. 322: 369-377, 1995; Barnes et al, Proc. Natl. Acad. Sci. USA 88: 5597-5601, 1991; Gillem et al, Arch Biochem Biophys 312: 59-66, 1994). A particularly preferred *E. coli* strain is strain C43(DE3) known to grow well while expressing a heterologous membrane protein in amounts which hold growth of commonly used strains. Thus, expression of CYP79B2 in the commonly used *E. coli* strain JM109 produced less than 0.5% of the CYP79B2 activity produced by strain C43(DE3). Expression in insect cells is also possible.

Investigations into the substrate specificity of CYP79D1, CYP79D2, CYP79E1, CYP79E2, CYP79A2, CYP79B2, CYP79B5 and CYP79F1 are carried out in E. coli spheroplasts reconstituted with sorghum NADPH-cytochrome P450 oxidoreductase in the presence of high amounts of lipids. L-α-dioleyl phosphatidyl choline and L-α-dilauroyl phosphatidyl choline are preferred lipids for the reconstitution. Both CYP79D1 and CYP79D2 are found to convert L-valine as well as L-isoleucine into their corresponding oximes. Both CYP79E1 and CYP79E2 are found to convert L-tyrosine into the corresponding oxime. CYP79A2 is found to convert L-phenylalanine into phenylacetaldoxime. CYP79B2 is found to convert tryptophan into indole-3-acetaldoxime. CYP79F1 is found to convert a chain-elongated methionine homologue into the corresponding aldoxime. Neither L-Leucine, L-phenylalanine nor L-tyrosine are metabolized by CYP79D1 or CYP79D2. Neither L-methionine, L-tryptophane nor L-tyrosine are metabolized by CYP79A2. Neither phenylalanine nor tyrosine are metabolized by CYP79B2. Neither L-tryptophane, L-phenylalanine nor L-tyrosine are metabolized by CYP79F1. D-Amino acids are not converted into oximes by CYP79D1, CYP79D2, CYP79E1 and CYP79E2. Depending on the nature of the substrate, substrate specificity may also be determined using intact *P. pastoris* cells or intact *E. coli* cells.

Nucleic acid compounds according to the invention consist of nucleotide residues independently selected from the group of the nucleotide residues G, A, T and C or the group of nucleotide residues G, A, U and C and are characterized by the formula R_A-R_B-R_C, wherein

- R_A, R_B and R_C designate component sequences; and
- R_B consists of at least 450 and preferably 600 or more nucleotide residues encoding amino acid component sequence R_2 as described above.

Knowledge of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, and SEQ ID NO: 4; SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, and SEQ ID NO: 12; SEQ ID NO: 39 and SEQ ID NO: 40; SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 70 and SEQ ID NO: 71; and SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 84 and SEQ ID NO: 85 can be used to accelerate the isolation and production of DNA coding for a P450 monooxygenase converting an aliphatic or aromatic amino acid or chain-elongated methionine homologue to the corresponding aldoxime which method comprises

(a) preparing a cDNA library from plant tissue expressing such a monooxygenase,

(b) using at least one oligonucleotide designed on the basis of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4; SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, or SEQ ID NO: 12; ; SEQ ID NO: 39 and SEQ ID NO: 40; SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 70 or SEQ ID NO: 71; or SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 84 or SEQ ID NO: 85 to amplify part of the P450 monooxygenase cDNA from the cDNA library,

(c) optionally using one or more oligonucleotides designed on the basis of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, or SEQ ID NO: 4; SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, or SEQ ID NO: 12; SEQ ID NO: 39 or SEQ ID NO: 40; SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 70 or SEQ ID NO: 71; or SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 84 or SEQ ID NO: 85 to amplify part of the P450 monooxygenase cDNA from the cDNA library in a nested PCR reaction,

(d) using the DNA obtained in steps (b) or (c) as a probe to screen the DNA library prepared from plant tissue expressing a P450 monooxygenase converting an aliphatic or aromatic amino acid or chain-elongated methionine homologue to the corresponding oxime, and

Expressed as transgenes DNA encoding P450 monooxygenases according to the present invention is particularly useful to modify the biosynthesis of glucosinolates or cyanogenic glucosides in plants. When the gene encoding a cytochrome P450 enzyme converting an aliphatic or aromatic amino acid into the corresponding oxime is expressed in an acyanogenic plant together with a cytochrome P450 enzyme belonging to the CYP71E family e.g. CYP71E1 from sorghum or preferably the corresponding homolog from cassava and a UDP-glucose cyanohydrin glucosyltransferase, the transgenic plant obtained will be cyanogenic. The introduction of the gene encoding a cytochrome P450 enzyme converting an aliphatic or aromatic amino acid or chain-elongated methionine homologue into the corresponding oxime into a plant species producing glucosinolates can be used to alter the glucosinolate production in said plants as observed by an alteration of the overall level or the content of individual glucosinolates in the transgenic plants selected. If the aliphatic or aromatic amino acid or chain-elongated methionine homologue that is the substrate of the introduced cytochrome P450 enzyme was not previously recognized as a substrate for other cytochrome P450s in that particular plant species, then a new glucosinolate is introduced in the transformed plant. Likewise, the introduction of the gene encoding a cytochrome P450 enzyme converting an aliphatic or aromatic amino acid into the corresponding oxime into a cyanogenic plant can be used to modify the overall level and profile of the preexisting cyanogenic glucosides and to introduce one or more additional cyanogenic glucosides in the plant.

Proper selection of promoters to provide constitutive, inducible or tissue specific expression of the genes provides means to obtain transgenic plants with desired disease or herbivor responses. Likewise, the content of glucosinolates or cyanogenic glucosides in plants may be modified or reduced using anti-sense or ribozyme technology using the same genes. Thus, it is a further aspect of the present invention to provide transgenic plants comprising stably integrated into their genome DNA comprising at least part of an open reading frame of a P450 monooxygenase according to the present invention converting an aliphatic or aromatic amino acid or chain-elongated methionine homologue to the corresponding oxime . Such plants can be produced by a method comprising

(a) introducing into a plant cell or tissue which can be regenerated to a complete plant, DNA comprising at least part of an open reading frame of a P450 monooxygenase according to the present invention converting an aliphatic or aromatic amino acid or chain-elongated methionine homologue to the corresponding oxime ; and

The gene specific fragment is labeled with digoxigenin-11-dUTP (Boehringer Mannheim, Germany) by PCR amplification and used as probe to screen the cassava cDNA library using the DIG system (Boehringer Mannheim, Germany). The probe is hybridized over night at 68°C in 5xSSC, 0.1% N-lauroylsarcosine, 0.02% SDS, 1% blocking reagent (Boehringer Mannheim, Germany). Prior to detection, filters are washed with 0.1 x SSC, 0.1% SDS at 65°C.

Example 2 - **CYP79D1 and CYP79D2, sequencing and southern blot analysis**
Using the probe obtained according to example 1 two equally abundant full-length clones are isolated from the cassava cDNA library. The clones have open reading frames encoding P450s of 61.2 and 61.3 kDa. These P450s are assigned CYP79D1 and CYP79D2 as the first two members of a new CYP79D subfamily.
Sequencing is performed using the Thermo Sequenase Fluorescent-labeled Primer cycle sequencing kit (7-deaza dGTP) (Amersham, Sweden) and an ALF-Express sequenator (Pharmacia, Sweden). Sequence computer analysis is performed using the programs from the GCG Wisconsin Sequence Analysis Package. The two cassava P450s are 85% identical and both share 54% identity to CYP79A1. P450s showing more than 40% but less than 55% sequence identity at the amino acid level are grouped in the same family but in different subfamilies.
The heme-binding motif in CYP79D1 and CYP79D2 is TFSTGRRGCVA (residues 470-480 of CYP79D1) and contains three amino acid substitutions compared to the consensus sequence PFGXGRRXCXG for A-type P450s (Durst et al, Drug Metabol Drug Interact 12: 189-206, 1995). The substitutions underlined are also found in CYP79A1 whereas the initial T in the CYP79D1 and CYP79D2 heme-binding motif is an S in CYP79A1, CYP79B1 and CYP79B2. Thus, the previously proposed existence of a heme binding sequence domain unique to the CYP79 family is contradicted. The other unique sequence domain PERH (residues 450-453 of CYP79D1), where H has been proposed to be specific for the CYP79 family is also found in CYP79D1 and CYP79D2.

To determine the copy number of *CYP79D1* and *CYP79D2*, a Southern Blot on genomic DNA from the cassava cultivar MCol22 is performed. Genomic DNA is purified from leaves of cassava cultivar Mcol22 as described by Chen et al in: *The Maize Handbook* (Freeling et al eds), Springer Verlag, NY, 1994. The DNA is further purified on Genomic-tip 100/G (Qiagen, Germany), digested with restriction enzymes and electrophoresed (10 µg

presence of *CYP79D1* or *CYP79D2* in zeocin resistant colonies is confirmed by PCR on the *P. pastoris* colonies.

Single colonies of *P. pastoris* are grown (28°C, 220 rpm) for approximately 22 h in 25 ml BMGY (1% yeast extract, 2% peptone, 0.1 M KP_i pH 6.0, 1.34% yeast nitrogen base, 4x 10^{-5}% biotin, 1% glycerol, 100 µg/ml zeocin). Cells are harvested (1500g, 10 min, RT) and inoculated in a 2 l baffled flask to OD_{600} of 0.5 in 300 ml of inducing medium, i.e. BMGY with 1% methanol instead of glycerol. The cultures are grown (28 °C, 300 rpm) for 28 h with addition of methanol to 0.5 % after 26 h. Cells are pelleted (3000g, 10 min, 4 °C) and washed once in buffer A (50 mM KP_i pH 7.9, 1 mM EDTA, 5% glycerol, 2 mM DTT, 1 mM phenylmethylsulfonyl fluoride) before being resuspended to OD_{600} of 130 in buffer A. An equal volume of acid-washed glass beads is added and the cells are broken by vortexing (8x 30 s, 4°C with intermediate cooling on ice). The lysate is centrifuged at 12000g (10 min, 4°C) to remove cell debris and the resulting supernatant recentrifuged at 165000g (1 h, 4 °C) to recover a microsomal pellet. Microsomes are resuspended in buffer A, stored at -80 °C and thawed on ice immediately before use.

CYP79D1 and CYP79D2 are functionally expressed in *P. pastoris* as evidenced by the ability of recombinant yeast cells to convert L-valine to the corresponding. No conversion took place using *P. pastoris* cells transformed with the vector only. The metabolic activity is measured in intact cells demonstrating that the endogenous *P. pastoris* reductase system is able to support electron donation to these plant P450s. SDS-PAGE of microsomes prepared from cells actively converting L-valine to val-oxime shows the presence of an additional polypeptide band migrating corresponding to a molecular mass of 62 kDa as expected from the *CYP79D1* cDNA clone.

With regard to CYP79D1 activity in intact *P. pastoris* cells the best results were obtained using growth in rich media and induction at OD 0.5 for 24-30 h. 15-30 nmol of microsomal CYP79D1 per liter culture are produced. The yield of microsomal CYP79D1 after 90 h of induction is 50% of that obtained after 24 h.

Example 4 - **Purification of recombinant CYP79D1**

All steps are carried out at 4 °C unless otherwise stated. CYP79D1 containing fractions are identified by carbon monoxide difference spectroscopy, SDS-PAGE and activity measurements.

This demonstrates that CYP79D1 is a fairly stable protein. Yeast cytochromes may interfere with the spectroscopy of crude extracts and hide a minor 420 nm peak and *P. pastoris* cytochrome oxidase had previously been reported to prevent P450 spectroscopy. In the present study, the expression level of CYP79D1 is high and the CO difference spectrum produced by cytochrome oxidase (maximum at 430 nm, minimum at 445) is visible as a shoulder on the 450 nm peak. The *P. pastoris* cytochrome oxidase binds to the DEAE column and accordingly is removed during P450 isolation. Upon culturing *P. pastoris* for extended periods (90 h), the content of cytochrome oxidase decreases permitting detection of lower amounts of P450 in microsomes. Finally, interfering cytochrome oxidase can be removed from P450 by TX-114 phase partitioning performed in borate buffer. Upon phase partitioning in borate, the P450s partition to the TX-114 poor phase, whereas *P. pastoris* cytochrome oxidase partitiones to the rich phase.
Purified CYP79D1 forms a type I substrate binding spectrum in the presence of L-valine corresponding to a 44 % shift from low spin to high spin state upon substrate binding.

Example 5 - **Determination of the catalytic activity**
Isolated, recombinant CYP79D1 is reconstituted and its catalytic activity determined *in vitro* using reaction mixtures with a total volume of 30 µl containing 2.5 pmol CYP79D1, 0.05 U NADPH P450-oxidoreductase (Benveniste et al, Biochem J 235: 365-373, 1986), 10.6 mM L-α-dioleyl phosphatidylcholine, 0.35 µCi [U-^{14}C]-L-amino acid (L-Val, L-Ile, L-Leu, L-Tyr or L-Phe; Amersham, Sweden), 1 mM NADPH, 0.1 M NaCl and 20 mM KP_i pH 7.9. In assays containing ^{14}C-L-valine or ^{14}C-L-isoleucine, different amounts of unlabeled L- and D-amino acids (0-6 mM) are added. After incubation for 10 minutes at 30 °C the products formed are extracted into 60 µl ethyl acetate and separated on TLC sheets (Merck Kieselgel 60F $_{254}$) using n-pentane/diethyl ether (50:50, v/v) or toluene/ethyl acetate (5:1, v/v) as eluents for aliphatic compounds and aromatic compounds, respectively. ^{14}C-labeled oximes are visualized and quantified using a STORM 840 phosphor imager (Molecular Dynamics, CA, USA). The activity of CYP79D1 is additionally measured in the presence of the inhibitors tetcyclasis, ABT and DPI under the same conditions as described above.
For *in vivo* activity assays 200 µl *P. pastoris* cells are pelleted and resuspended in 100 µl 50 mM Tricine pH 7.9 and 0.35 µCi [U-^{14}C]-L-valine or L-isoleucine. After incubation for 30 minutes at 30°C the cells are extracted with ethyl acetate and the products formed are analyzed as above.

Example 6 - **N-terminal sequencing of CYP79D1**

Isolated recombinant CYP79D1 is subjected to SDS-PAGE and the protein transferred to ProBlott membranes (Applied Biosystems, CA, USA) as described in Kahn et al, J. Biol. Chem 271: 32944-32950, 1996. The Coomassie Brilliant Blue-stained protein band is excised from the membrane and subjected to sequencing on an Applied Biosystems model 470A sequenator equipped with an on-line model 120A phenylthiohydantoin amino acid analyzer. Asn glycosylation is detected as the lack of an Asn signal in the predicted Edman degradation cycle.

The fractions that produce CO spectra and contain CYP79D1 activity always produce two distinct closely migrating polypeptide bands upon SDS-PAGE. N-terminal amino acid sequencing identifies both bands as derived from CYP79D1. The initial methionine is removed by the yeast processing system. Sequencing of the first 15 residues of the upper band demonstrates glycosylation of both asparagines present, whereas the lower band only is glycosylated at the first asparagine. The different glycosylation pattern explains the presence of two bands. Glycosylation at the N-terminal part of CYP79D1 is in agreement with the localization of the N-terminal in the lumen of the endoplasmatic reticulum accessible for the glycosylation machinery. It is unknown, whether native CYP79D1 is glycosylated in cassava. However, CYP79A1 purified from sorghum seedlings is not glycosylated as documented by amino acid sequencing of the N-terminal fragment (15) and only few reports exist of microsomal P450 glycosylation. The observed glycosylation of recombinant CYP79D1 upon expression in *P. pastoris* is thought to reflect expression in a yeast system.

Example 7 - **Primers used in examples 8 and 9**

Primer Designation	**Nucleotide sequence** [a]	**SEQ ID NO:**
1F [b]	GCGGAATTCGAYAAYCCIWSIAAYGC	13
1R [b]	GCGGATCCGCIACRTGIGGIAHRTTRAA	14

[e] Covers a sequence that is identical in the two clones #1 and #2.

[f] Covers a sequence that is specific for either of the two clones #1 and #2.

[g] A specific primer for the 5'UTR in #1.

[h] The star indicates a stop codon.

Example 8 - **cDNA cloning of *Triglochin maritima* CYP79 genes**

PCR approach to generate cDNA fragments of a CYP79 homologue in *T. maritima*

A unidirectional plasmid cDNA library is made by In Vitrogen (Carlsbad, CA) from flowers and fruits (schizocarp) of *T. maritima*, using the expression vector pcDNA2.1 which contains the *lacZ* promoter. Plant material is collected at Aflandshage on Southern Amager, at the coast of Øresund, frozen directly in liquid N_2 and stored at -80°C.

Degenerate PCR primers are designed based on conserved amino acid sequences in CYP79A1 derived from *S. bicolor* - GenEMBL U32624, CYP79B1 from *Sinapis alba* - GenEMBL AF069494, CYP79B2 from *Arabidopsis thaliana* - GenEMBL, and a PCR fragment of CYP79D1 from *Manihot esculenta* - GenEMBL AF140613. Two rounds of PCR amplification reactions in a total volume of 50 µl are carried out using 100 pmol of each primer, 5% dimethyl sulfoxide, 200 µM dNTPs and 2.5 units *Taq* DNA polymerase in PCR buffer (50 mM KCl, 10 mM Tris-HCl pH 8.8, 1.5 mM $MgCl_2$, 0.1% Triton X-100). Thermal cycling parameters are 2 min at 95°C, 30 × (5 sec at 95°C , 30 sec at 45°C, 45 sec at 72°C) and finally 5 min at 72°C. The first PCR reaction is performed using primers 1F and 1R (Example 7) on 100 ng template DNA prepared from the cDNA library or genomic DNA prepared using the Nucleon Phytopure Plant DNA Extraction Kit (Amersham). The PCR products are purified using QIAquick PCR Purification Kit (Qiagen), eluted in 30 µl 10 mM Tris-HCl pH 8.5, and used as template (1 µl) for the second round of PCR reactions carried out using PCR fragments derived from both cDNA and genomic DNA and using the two degenerate primers 2F and 2R (Example 7). An aliquot (5 µl) of the PCR reaction is applied to a 1.5% agarose/TBE gel and a band of the expected size of about 200 bp is observed using both cDNA and genomic DNA as template. The rest of the PCR reaction is purified using QIAquick PCR Purification Kit and eluted in 30 µl 10 mM Tris-HCl pH 8.5. The purified PCR fragments (5µl) are digested with *Eco*RI and *Bam*HI, excised from a 1.5% agarose/TBE gel, purified using QIAEX II Agarose Gel Extraction kit (Qiagen) and ligated into an *Eco*RI- and *Bam*HI-digested pBluescript II SK vector (Stratagene). Seven clones derived from the cDNA library and three clones derived from genomic DNA are sequenced

a consequence, the PCR fragment cloned with 4R#1 and 3R is used as a template to generate a digoxigenin-11-dUTP labeled probe (TRI2) using primers 5F#1 and 5R#1 (Example 7). Using the same conditions as above, TRI2 partly covering the 5' untranslated region (UTR) and 5' end of the open reading frame of clone #1 is used to screen the pcDNA2.1 library together with the TRI1 probe. The first lifts are hybridized with TRI2 and the second with TRI1. Two individual cDNA clones with exactly the same length as the PCR fragment are isolated after screening 1.000.000 colonies.

Results

Based on a sequence alignment of CYP79A1 and putative N-hydroxylases belonging to the CYP79 family, four degenerate oligonucleotide primers covering two CYP79 specific regions are designed (1F, 2F, 1R, 2R described in Example 7) and used in nested PCR reactions with genomic DNA as well as cDNA made from flowers and fruits of *Triglochin maritima* as templates. A PCR fragment of the expected size, i.e. approximately 200 bp, and showing 62 to 70% identity to CYP79 sequences at the amino acid level is amplified from both templates, cloned and further used to screen the cDNA library. Two cDNA clones, denoted #1 and #2, are isolated and verified by sequence comparison to share high sequence identity to the CYP79 family. Using clone specific PCR primers, a full-length clone corresponding to #1 is isolated. The open reading frame encodes a protein with a molecular mass of 60.8 kDa. A comparison of the full-length sequence of clone #1 with that of clone #2 reveals that clone #2 is 6 bp shorter at the 5´ end but contains a methionine codon not found in clone #1 at a position corresponding to amino acid residue 26 specified by clone #1. The sequence surrounding this methionine codon does not fit the general context sequence for a start codon in a monocotyledonous plant. Most likely, clone #2 thus lacks 6 bp to be full-length.

The cytochrome P450s encoded by clones #1 and #2 show 44 to 48% identity to already known members of the CYP79 family (see Table below) and accordingly are identified as the first two members of the new subfamily CYP79E and assigned CYP79E1 (SEQ ID NO: 9) and CYP79E2 (SEQ ID NO: 11). The sequence identity between CYP79E1 and CYP79E2 is 94%.

truncated constructs are made using primers 6F#1($\Delta(1\text{-}31)_{17\alpha(8aa)}$) and 6R#1 or primers 6F#1($\Delta(1\text{-}52)_{2E1(10aa)}$) and 6R#1 (Example 7). Construct CYP79E1 $\Delta(1\text{-}31)_{17\alpha(8aa)}$ encodes a truncated form of CYP79E1 in which 31 codons of the native 5' sequence are replaced by 8 AT-enriched codons of P45017 α (Halkier et al, Arch. Biochem. Biophys. 322: 369-377, 1995; Barnes et al, Proc. Natl. Acad. Sci. USA 88: 5597-5601, 1991); in construct CYP79E1$\Delta(1\text{-}52)_{2E1(10aa)}$ the first 52 codons of the native 5´sequence are replaced by 10 AT-enriched codons of P4502E1and silent mutations are introduced in codons 53 and 55.The PCR fragments are digested with *Nde*I and *Hin*dIII and ligated into *Nde*I- and *Hin*dIII-digested pSP19g10L expression vector (Barnes, Methods Enzymol. 272: 3-14, 1996). The unique restriction sites *Nco*I and *Pml*I are used to replace the middle part of the PCR clones (1045 bp) with the analogous fragment from the cDNA clone. The remaining portions of the constructs deriving from PCR, are sequenced to exclude PCR errors.

Because the CYP79E2 clone is isolated in frame with the first 24 codons of the *lacZ* gene in the vector pcDNA2.1, this clone is tested as a fourth expression construct designated CYP79E2$_{lacZ(24aa)}$. For comparison, an equivalent fifth construct CYP79E1 $\Delta(1\text{-}2)_{lacZ(24aa)}$ is also prepared.

All constructs contain the original stop sequence TAAT found in most highly expressed *E. coli* genes. All constructs using the vector pSP19g10L have their 3´UTR removed, because inclusion of the 3´UTR has been reported to prevent or reduce expression of some genes. In constructs based on pcDNA2.1, the 3´UTR is retained.

Expression in *E. coli*

All expression constructs are transformed into the *E. coli* strains JM109 (Stratagene) and XL-1 blue (Stratagene). In all cases, the JM109 strain turns out to be most efficient.

CYP79E1 and CYP79E2 contain 19 and 17 AGA or AGG arginine codons which are rare in *E. coli* genes. A strong positive correlation between the occurrence of codons and tRNA content has been established. Accordingly, the native and $\Delta(1\text{-}52)_{2E1(10aa)}$ constructs of clone #1 as well as the construct of clone #2 are co-transformed with pSBET (Schenk et al, BioTechniques 19: 196-200, 1995) encoding a tRNA gene for rare arginine codons, into JM109. Single colonies are grown overnight in LB medium (50 µg/ml ampicillin, 37°C, 225

Before TLC application the sample is extracted with ethyl acetate. During this step the surplus of radiolabeled tyrosine remains in the aqueous phase thus preventing overexposure at the origin. The total ethyl acetate phase is applied to the TLC plate. In some experiments, inevitable carry-over of small amounts of the aqueous phase results in the appearance of a tyrosine band at the origin. Unlabeled reference compounds (*p*-hydroxyphenylacetaldoxime, *p*-hydroxyphenylacetonitrile and *p*-hydroxybenzaldehyde) are prestreaked on the TLC plates to permit visual detection under ultraviolet light.

Carbon monoxide binding spectra using intact *E. coli* cells show the absorption maximum at 450 nm diagnostic for formation of functional cytochrome P450 with the following three constructs: CYP79E1$_{na}$, CYP79E1$\Delta(1\text{-}52)_{2E1(10aa)}$, and CYP79E2$_{lacZ(24aa)}$. The spectra are obtained without and with co-transformation of pSBET but in all cases the cytochrome P450 content turns out to be too low to permit quantification. To obtain an accurate determination, the cytochrome P450s are enriched by isolation of *E. coli* spheroblasts followed by temperature-induced Triton X-114 phase partitioning (Werck-Reichart et al, Anal. Biochem. 197: 125-131, 1991; Halkier et al, Arch. Biochem. Biophys. **322:** 369-377, 1995). The highest expression level (in JM109 cells after 48 hours) of 56 nmol/l culture is obtained using CYP79E2$_{lacZ(24aa)}$. This level is comparable to the expression level of 62 nmol/l culture obtained with *S. bicolor* construct CYP79A1$\Delta(1\text{-}33)_{17\alpha(8aa)}$ (Halkier et al, Arch. Biochem. Biophys. **322:** 369-377, 1995) included as a positive control. CYP79E1 $\Delta 1\text{-}31)_{17\alpha(8aa)}$ with a modified P45017α N-terminal and the empty vector do not reveal any detectable spectrum.

Example 10 - **Reconstitution of CYP79E with CYP71E1**

Reconstitution of the membrane associated pathway of cyanogenic glucoside synthesis resulting in the formation of *p*-hydroxymandelonitrile, the aglycon of dhurrin (seen as *p*-hydroxybenzaldehyde *in vitro*) is achieved using enzymes from the two species *S. bicolor* and *Triglochin maritima*. In reconstitution experiments including tyrosine, NADPH, NADPH-cytochrome P450 oxidoreductase, CYP71E1 and CYP79E1 or CYP79E2, considerable amounts of *p*-hydroxyphenylacetonitrile and *p*-hydroxybenzaldehyde accumulate.

Example 11 - **Primers used in examples 12 and 13**

The following PCR primers are designed on the basis of the genomic *Arabidopsis thaliana* L. cv. Columbia sequence of CYP79A2 found to be contained in GenBank Accession

digested pYX223 (R&D Systems), and inserts of 10 clones derived from two nested PCR reactions are sequenced.
Sequencing is performed using the Thermo Sequence Fluorescent-labelled Primer cycle sequencing kit (7-deaza dGTP) from Amersham Pharmacia Biotech and analyzed on an ALF-Express DNA Sequencer (Amersham Pharmacia Biotech). Sequence computer analysis is done with programs of the GCG Wisconsin Sequence Analysis Package. The GAP program is used with a gap creation penalty of 8 and a gap extension penalty of 2 to compare pairs of sequences. The splice site prediction is done using NetPlantGene.

CYP79A2 is one of several *CYP79* homologues identified in the genome of *A. thaliana*. According to computer-aided splice site prediction it contains one intron, which is characteristic for A-type cytochromes P450. While it is the only intron in *CYP79A2* other members of the *CYP79* family have one or two additional introns. The sequence of the full-length *CYP79A2* cDNA confirms the splice site prediction. The reading frame of the *CYP79A2* cDNA has two potential ATG start codons, one positioned 15 bp downstream of a stop codon in the 5'untranslated region and another one 15 bp further downstream. The cDNA starting with the second ATG codon is for all further studies. This cDNA encodes a protein of 523 amino acids which has 64% similarity and 53% identity to CYP79A1 involved in the biosynthesis of the cyanogenic glucoside dhurrin.

Example 13 - **CYP79A2 *E. coli* expression constructs**

Expression constructs are derived from a *CYP79A2* cDNA obtained by fusion of the two exons amplified from genomic DNA of *Arabidopsis thaliana* L. The two exons are amplified by PCR with the primers A2F2 and A2R3 for exon 1 and A2F3 and A2R2 for exon2, respectively and using 1.25 units *Pwo* polymerase (Roche Molecular Biochemicals) and 4 mg template DNA. PCR reactions are set up in a total volume of 50 µl in *Pwo* polymerase PCR buffer with 2 mM $MgSO_4$ (Roche Molecular Biochemicals) supplemented with 200 µM dNTPs, 50 pmol of each primer, and 5 (v/v) % DMSO. After incubation of the reactions at 94°C for 3 minutes, 30 PCR cycles of 20 seconds at 94°C, 10 seconds at 60°C, and 30 seconds at 72°C are run. After digestion of the PCR fragments with *Eco*RI (exon 1) and *Hind*III (exon 2), the blunt ends generated with primers A2R3 and A2F3 and *Pwo* polymerase are phosphorylated with T4 polynucleotide kinase (New England Biolabs). The

different *CYP79A2* cDNAs are excised from pYX223 by digestion with *Nde*I and *Hind*III and ligated into *Nde*I/*Hind*III-digested pSP19g10L.

Example 14 - **CYP79A2 Expression in *E. coli***

E. coli cells of strain JM109 transformed with the expression constructs described in Example 13 are grown overnight in LB medium supplemented with 100 μg ml^{-1} ampicillin and used to inoculate 100 ml modified TB medium containing 50 μg ml^{-1} ampicillin, 1 mM thiamine, 75 μg ml^{-1} δ-aminolevulinic acid, and 1 mM isopropyl-β-D-thiogalactoside. The cells are grown at 28°C for 65 hours at 125 rpm. Cells from 75 ml culture are pelleted and resuspended in buffer composed of 0.1 M Tris HCl pH 7.6, 0.5 mM EDTA, 250 mM sucrose, and 250 μM phenylmethylsulfonyl fluoride. Lysozyme is added to a final concentration of 100 μg ml^{-1}. After incubation for 30 minutes at 4°C, magnesium acetate is added to a final concentration of 10 mM. Spheroplasts are pelleted, resuspended in 5 ml buffer composed of 10 mM Tris HCl pH 7.5, 14 mM magnesium acetate, and 60 mM potassium acetate pH 7.4 and homogenized in a Potter-Elvehjem. After DNAse and RNAse treatment, glycerol is added to a final concentration of 29%. Temperature-induced Triton X-114 phase partitioning is performed as described in Halkier et al, Arch Biochem Biophys 322: 369-377, 1995. The Triton X-114 rich phase is analyzed by SDS-PAGE.

Fe^{2+}·CO vs. Fe^{2+} difference spectroscopy (Omura et al, J Biol Chem 239: 2370-2378, 1964) is performed on 100 μl *E. coli* spheroplasts resuspended in 900 μl of buffer containing 50 mM KP_i pH 7.5, 2 mM EDTA, 20% (v/v) glycerol, 0.2% (v/v) Triton X-100, and a few grains of sodium dithionite. The suspension is distributed between two cuvettes and a baseline is recorded between 400 and 500 nm on a SLM Aminco DW-2000 TM spectrophotometer (SLM Instruments, Urbana, IL). The sample cuvette is flushed with CO for 1 min and the difference spectrum is recorded. The amount of functional cytochrome P450 is estimated based on an absorption coefficient of 91 l $mmol^{-1}$ cm^{-1}.

The activity of CYP79A2 is measured in *E. coli* spheroplasts reconstituted with NADPH:cytochrome P450 oxidoreductase purified from *Sorghum bicolor* (L.) Moench as described in Sibbesen et al, J Biol Chem 270: 3506-3511, 1995. In a typical enzyme assay, 5 μl spheroplasts and 4 μl NADPH:cytochrome P450 reductase (equivalent to 0.04 units defined as 1 μmol cytochrome c min^{-1}) are incubated with 3.3 μM L-[U-^{14}C]phenylalanine

functional cytochrome P450. A peak at 415 nm is found for all spheroplast preparations. This peak may arise from *E. coli* derived heme protein, unattached heme groups produced in the presence of δ-aminolevulinic acid in the medium, or cytochrome P450 in a non-functional conformation. Based on the peak at 452 nm, the expression level of 'chimeric' CYP79A2 is estimated to be 50 nmol cytochrome P450 (l culture) $^{-1}$. When incubated with L-[^{14}C]phenylalanine, spheroplasts of *E. coli* transformed with the 'native', the 'truncated-modified', or the 'chimeric' CYP79A2 expression construct and reconstituted with the purified NADPH:cytochrome P450 oxidoreductase from *S. bicolor* produce two radiolabelled compounds which comigrate with the (*E*)- and (*Z*)-isomers of phenylacetaldoxime in thin layer chromatography. These products are not detected in assay mixtures containing *E. coli* spheroplasts harbouring either the 'modified' CYP79A2 expression construct or the empty vector. GC-MS analysis shows that two compounds with identical fragmentation patterns are present in the reaction mixture with 'chimeric' CYP79A2, but not in the control reaction. The retention times and the fragmentation pattern identify these compounds as the (*E*)- and (*Z*)-isomers of phenylacetaldoxime. Administration of L-[^{14}C]tyrosine, L-[^{14}C]methionine, or L-[^{3}H]tryptophan to spheroplasts of *E. coli* expressing the 'native' or the 'chimeric' CYP79A2 does not result in production of detectable amounts of the respective aldoximes. The ability of CYP79A2 to metabolize DL-homophenylalanine is investigated in spheroplasts of *E. coli* expressing 'chimeric' CYP79A2. GC-MS analysis of the reaction mixture shows the absence of detectable amounts of the homophenylalanine-derived aldoxime. A K_m value of 6.7 µmol l^{-1} and a V_{max} value of 16.6 pmol min^{-1} (mg protein)$^{-1}$ are determined for CYP79A2 using spheroplasts of *E. coli* expressing 'native' CYP79A2 with L-[^{14}C]phenylalanine as the substrate. As no CO spectrum is obtained with 'native' CYP79A2, it is not possible to estimate the amount of functional 'native' CYP79A2. However, based on the expression level of functional 'chimeric' CYP79A2, a turnover number of 0.24 min $^{-1}$ for 'native' CYP79A2 can be estimated.

The substrate specificity of CYP79A2 seems to be rather narrow as neither L-tyrosine, DL-homophenylalanine, L-tryptophan nor L-methionine are metabolized by the enzyme. The high substrate specificity is in agreement with results obtained with CYP79 homologues involved in the biosynthesis of cyanogenic glucosides, The activity of recombinant CYP79A2 is strongly dependent on the pH of the reaction mixture and, to a lesser extent, on several other factors. Compared to the activity at pH 7.5, the activity of 'chimeric' CYP79A2 is 25% at pH 6, 50% at pH 6.5, 80% at pH 7.0, and 70% at pH 7.9. Addition of

loaded on a DEAE Sephadex A-25 column equilibrated as follows: 25 mg DEAE Sephadex A-25 are swollen overnight in 1 ml 0.5 M acetate buffer pH 5, packed into a 5 ml pipette tip, and washed with 1 ml water. The plant extract is loaded, and the column is washed with 2 ml 70% (v/v) methanol, 2 ml water, and 0.5 ml 0.02 M acetate buffer pH 5. *Helix pomatia* sulfatase (Type H-1, Sigma; 0.1 ml, 2.5 mg ml^{-1} in 0.02 M acetate buffer pH 5) is applied, and the column is left at room temperature for 16 hours. Elution is carried out with 2 ml water. The eluate is dried *in vacuo*, the residue dissolved in 150 µl water, and 100 µl are subjected to HPLC on a Shimadzu LC-10A T *vp* equipped with a Supelcosil LC-ABZ 59142 C_{18} column (25 cm x 4.6 mm, 5 mm; Supelco) and a SPD-M10AVP photodiode array detector (Shimadzu). The flow rate is 1 ml min^{-1}. Elution with water for 2 minutes is followed by elution with a linear gradient from 0 to 60% methanol in water (48 minutes), a linear gradient from 60 to 100% methanol in water (3 minutes) and with 100% methanol (3 minutes). The assignment of peaks is based on retention times and UV spectra compared to standard compounds. Glucosinolates are quantified in relation to the internal standard and by use of the response factors as described by Buchner (1987) In: Glucosinolates in rapeseed: Analytical aspects, Wathelet, (ed.), Martinus Nijhoff Publishers, pp 50-58 and Haughn et al, Plant Physiol 97: 217-226,1991. In the analysis of rosette leaves, the term 'total glucosinolate content' refers to the molar amount of the five major glucosinolates (4-methylsulfinylbutylglucosinolate, 4-methylthiobutylglucosinolate, 8-methylsulfinyloctylglucosinolate, indol-3-ylmethylglucosinolate, and 4-methoxyindol-3-ylglucosinolate) which account for 85% of the glucosinolate content in rosette leaves of wild-type *A. thaliana* and benzylglucosinolate. The glucosinolate content of transgenic seeds harvested from T1 plants #10, #13, and #14 is analyzed and compared with the glucosinolate content of wild-type seeds. Twelve to thirty milligrams of seeds are extracted and subjected to HPLC analysis as described above with the exception that lyophilization of the tissue is omitted. In this analysis of seeds, the term 'total glucosinolate content' refers to the molar amount of the ten major glucosinolates (3-hydroxypropylglucosinolate, 4-hydroxybutylglucosinolate, 4-methylsulfinylbutylglucosinolate, 4-methylthiobutylglucosinolate, 8-methylsulfinyloctylglucosinolate, 7-methylthioheptylglucosinolate, 8-methylthiooctylglucosinolate, indol-3-ylmethylglucosinolate, 3-benzoyloxypropylglucosinolate, 4-benzoyloxybutylglucosinolate) which account for more than 90% of the glucosinolate content in seeds of wild-type *A. thaliana* and benzylglucosinolate.

presented, however, indicate that aldoxime formation from aromatic amino acids is dependent on cytochrome P450 enzymes in members of the *Brassicaceae* as well as in other families.

Example 16 - **Expression analysis of CYP79A2 by histochemical GUS assay**

The CYP79A2 promoter is studied in transgenic *A. thaliana* transformed with a construct containing the CYP79A2 promoter in front of the GUS-intron DNA sequence. A genomic clone containing the CYP79A2 gene is isolated from the EMBL3 genomic library (*A. thaliana* cv. Columbia). A *SacI/XmaI* fragment (SEQ ID NO: 15) consisting of 2.5 kB upstream sequence and 120 bp CYP79A2 coding region is excised from the DNA of the positive phage. The fragment is inserted into pPZP111 in frame with the *XbaI/SalI* fragment of pVictor IV S GiN (Danisco Biotechnology, Denmark) containing the GUS-intron sequence and the 35S terminator. The fusion between the two fragments is made by a 17 bp linker. The resulting transcript encodes a fusion protein consisting of the CYP79A2 membrane anchor fused to the GUS protein.

Transformants of different developmental stages are analyzed by histochemical GUS assays. Intense staining is observed in the veins of the hypocotyl and the petioles of ten days old plants. No staining is seen in the cotelydones and leaves except of the hydathodes where intense staining is observed. In three weeks old plants the veins of the leaves are stained with moderate intensity while intense coloration is observed in the hydathodes. No staining is found in roots of ten days and three weeks old plants. In five weeks old plants no GUS activity is detected.

Example 17 - ***Arabidopsis*** **plants and primers used in examples 18, 19, 21, and 22**

Arabidopsis cv. Columbia is used for all experiments. Plants are grown in a controlled-environment Arabidopsis Chamber (Percival AR-60 I, Boone, Iowa, USA) at a photosynthetic flu: of 100-120 µmol photons m^{-2} sec^{-1}, at 20°C and 70% relative humidity. The photoperiod is 12 hours for plants used for transformation and 8 hours for plants used for biochemical analysis.

specific EST3 primer a 255 bp fragment of the missing 5' end is amplified and subsequently cloned by use of an *Eco*R I site in the amplified vector sequence and a *Bam*H I site introduced by primer EST3. This fragment is used as template to amplify a Digoxigenin-11-dUTP (DIG, Boehringer Mannheim) labelled probe (DIG1) by PCR with primers EST6 and EST7A. The λPRL2 library is screened with the DIG1 probe according to the manufacturer's instructions (Boehringer Mannheim) hybridization occurring overnight at 68 °C in 5x SSC, 0.1% N-lauroyl sarcosin, 0.02% SDS, 1.2% (w/v) blocking reagent (Boehringer Mannheim) and stringency washes being performed two times for 15 minutes at 65 °C, 0.1x SSC, 0.1 % SDS. Detection of positive plaques is done by chemiluminescent detection with nitro blue tetrazolium according to the manufacturer's instructions (Boehringer Mannheim). Screening of the λPRL2 library with the 255 bp PCR fragment as a probe (DIG1) results in the isolation of a full length cDNA clone encoding CYP79B2.

EST T42902 is identified based on homology to the S. bicolor CYP79A1 sequence. A 240 bp PCR fragment is amplified with primers EST1 and EST2 using EST T42902 from the Arabidopsis Biological Research Center at OHIO State University as template. This PCR fragment is labelled with Digoxigenin-11-dUTP (DIG, Boehringer Mannheim) and used as probe to screen a lambda ZAP II cDNA library from *Brassica napus* leaves (Clontech Lab., Inc.). The library is screened with the DIG probe according to the manufacturers instructions, hybridizations occurring overnight at 68°C in 5x SSC, 0.1% N-lauryl sarcosin, 0.02% SDS, 1.2% (w/v) blocking reagent (Boehringer Mannheim) and stringency washes being performed two times for 15 minutes at 65°C, 0.1x SSC, 0.1% SDS. Positive plaques are detected by chemiluminescent detection with nitro tetrazolium according to the manufacturers instruction (Boehringer Mannheim). Screening of the library results in the isolation of a full length cDNA clone encoding CYP79B5.

The sequence reactions are performed using the Thermo Sequence Fluorescent-labelled Primer cycle sequencing kit (Amersham) and analyzed on an ALF-express automated sequenator (Pharmacia). Sequence computer analysis and alignments are produced with programs in the Wisconsin Sequence Analysis Package.

For Southern Blot Analysis genomic DNA is isolated from Arabidopsis leaves with the Nucleon PhytoPure Plant DNA extraction kit (Amersham). 10 µg of DNA are digested with *Bam*H I, *Xba* I, *Ssp* I, *Eco*R I or *Eco*R V and fractionated by gel electrophoresis on a 0.8% agarose gel. Southern blot analysis is performed with the Digoxigenin labelled probe DIG1 and washed under high stringency conditions (68°C, 0.1x SSC, 0.1% SDS, 2x 15 minutes). Bands are visualized by chemiluminescent detection with CDP-Star™ (Tropix Inc.).

highest level of expression is found in roots, the lowest level in stem leaves; approximately equal amounts are found in rosette leaves, stems and flowers. The level of CYP79B2 messenger RNA in roots is approximately 3-4 fold higher than the level found in rosette leaves. A two-fold induction detectable within 15 minutes after wounding is seen in rosette leaves after 2 hours. Said increase is in agreement with CYP79B2 being involved in indoleglucosinolate biosynthesis.

Example 19 - **CYP79B2 *E. coli* expression constructs and activity measurement**

PCR with the 5´ 'native' sense primer or the 5´ 'bovine' sense primer against the 3´ 'end' antisense primer are used to generate the constructs 'native' and ' $\Delta(1\text{-}9)_{bov}$', respectively, for expression. Using the *Aat* II and *Nde* I restriction sites introduced by the primers, the PCR fragments are cloned into an *Aat* II /*Nde* I digested pSP19g10L vector (Barnes, Meth. Enzymol. 272: 3-14, 1996) and sequenced to exclude PCR errors.
The native construct consists of the unmodified coding region of CYP79B2, whereas the $\Delta(1\text{-}9)_{bov}$ construct is truncated by 9 amino acids, in addition to having the first eight codons replaced by the first eight codons of bovine P45017 α (17). The bovine modification has been shown to result in high level expression of cytochrome P450s in *E. coli*. Both constructs carry the modified stop sequence of TAA T to increase translational stop efficiency (Tate et al, Biochem. 31, 2443-2450, 1992).

The activity of CYP79B2 is measured by reconstituting spheroplasts from *E. coli* expressing CYP79B2 with purified NADPH:cytochrome P450 reductase from *Sorghum bicolor* (L.) Moench. The *S. bicolor* NADPH:cytochrome P450 reductase is purified as described by Sibbesen et al, J. Biol. Chem. 270: 3506-3511, 1995. The reaction is started by addition of 5 µl of *E. coli* spheroplasts to a 45 µl reaction mixture containing 100 mM Tricine pH 7.9, 10 µg/µl DLPC (dilaurylphosphatidylcholine) sonicated for 2x 10 seconds, 4 mM NADPH, 3 mM reduced glutathiona (GSH), 5 µl [3-^{14}C]tryptophan (0.1 µCi, specific activity 56.5 mCi/mmol) and 1 U/µl purified NADPH:cytochrome P450 reductase. The reaction is incubated at 34°C for 30 minutes, extracted two times with ethyl acetate and the ethyl acetate phase is analyzed by TLC using toluen:ethyl acetate 5:1 as eluent. Radiolabelled bands are visualized on a Storm 840 phosphorimager (Molecular Dynamics) and quantified with ImageQuant analysis software (Molecular Dynamics). Substrate specificity is investigated by substituting the ^{14}C-labelled tryptophan with ^{14}C-labelled tyrosine or phenylalanine.

is determined to be 21 µM and V_{max} is determined to be 97.2 pmol/h/µl spheroplast. No oxime producing activity is detected when radiolabelled phenylalanine or tyrosine are administered to reaction mixtures containing recombinant CYP79B2. This indicates that CYP79B2 is specific for tryptophan.

CO-difference spectra of spheroplasts or of the rich phase of a Triton X-114 temperature-induced phase partitioning from the spheroplasts does not show a characteristic peak at 450 nm. Furthermore, when spheroplasts or the Triton X-114 rich phase thereof are separated on an SDS-polyacrylamide gel and stained with Coomassie Brilliant Blue a new band of approximately 60kD is visible. This indicates that very little recombinant CYP79B2 is produced and that CYP79B2 is highly active.

Plasma membrane enzyme systems in Chinese cabbage and Arabidopsis have previously been shown to catalyze the formation of IAOX from tryptophan via a peroxidase-like enzyme (TrpOxE). The conversion is stimulated by H_2O_2 and in certain cases by $MnCl_2$ and 2,4-dichlorophenol. Addition of 100 mM H_2O_2, 1 mM $MnCl_2$ or 800 µM 2,4-dichlorophenol to the CYP79B2 reconstitution assays inhibits the activity by 96%, 34% and 72%, respectively, and by 99% when combined. This shows that the two systems are not identical and that the TrpOxE activity is clearly distinctg from CYP79B2. Moreover, a non-enzymatic reaction mixture containing 100 mM H_2O_2, 1 mM $MnCl_2$ and 800 µM 2,4-dichlorophenol in 50 mM Tricine buffer, pH 8.0 is able to catalyze the conversion of tryptophan to a compound co-migrating with IAOX at a conversion rate of approximately 0.7% of that seen for CYP79B2. This indicates that non-enzymatic conversion of tryptophan to IAOX can occur under oxidative conditions.

Example 21 - **Sense and antisense expression of CYP79B2 in Arabidopsis thaliana**

CYP79B2 cDNA is cloned in sense and antisense direction behind the cauliflower mosaic virus 35S (CaMV35S) promoter using the primers CYP79B2.2, B2SB, B2AF, and B2AB. The native full-length CYP79B2 cDNA is amplified by PCR using the primer pair CYP79B2.2 / B2SB (sense construct) and B2AF / B2AB (antisense construct). The PCR product for the sense construct is cloned into *Eco*R I/*Xba* I digested pRT101 (Töpfer et al, Nucleic Acid Res 15: 5890, 1987) and sequenced. The PCR product for the antisense construct is cloned into *Eco*R I/*Xho* I digested pBluescript (Stratagene), excised by digestion with *Eco*R I and *Kpn* I, and ligated into *Eco*R I/*Kpn* I digested pRT101 and sequenced. The sense and antisense expression cassettes are excised from pRT101 by *Pst* I digestion and

tyrosine to *p*-hydroxyphenylacetaldoxime, resulted in dwarfed plants with high content of the tyrosine-derived *p*-hydroxybenzylglucosinolate. The *p*-hydroxyphenylacetaldoxime produced by CYP79A1 was very efficiently channelled into *p*-hydroxybenzylglucosinolate. A similar efficient channelling of IAOX into indoleglucosinolates might also occur in the Arabidopsis overexpressing CYP79B2. However, it cannot be excluded that the dwarf phenotype is due to increased levels of IAA produced from IAOX, or from indole-3-acetonitrile generated from degradation of the increased level of indoleglucosinolates.

HPLC analyses of glucosinolate profiles of the T_1 generation of transgenic Arabidopsis shows that plants overexpressing CYP79B2 accumulate higher quantities of indoleglucosinolates than control plants transformed with empty vector. The levels of the two most abundant indoleglucosinolates glucobrassicin and 4-methoxyglucobrassicin are increased by approximately five fold and two-fold, respectively, whereas the level of neoglucobrassicin is not increased significantly. The total glucosinolate content is increased due to the higher levels of indoleglucosinolates, but the levels of aliphatic and aromatic (i.e. non-indole-) glucosinolates are not affected. In the antisense plants the level of indoleglucosinolates is not reduced compared to control plants. A possible explanation is that the antisense constructs used provide an insufficient means of downregulating CYP79B2. Alternatively, CYP79B3, which based on homology is likely to catalyze the same reaction, compensate the downregulation of indoleglucosinolates.

Example 22 - **Expression analysis of CYP79B2 by histochemical GUS assay**

Using the DIG system (Boehringer) an Arabidopsis ecotype Columbia EMBL3 genomic library is screened with a 505 bp Digoxigenin-11-dUTP labelled probe annealing to the 5' end of the CYP79B2 gene. Hybridization of the probe is done at 65°C in 5x SSC, 0.1% N-lauroylsarcosine, 0.02% SDS, and 1% blocking reagent. Filters are washed in 0.1x SSC, 0.1% SDS at 65°C prior to detection. Phage DNA from the positive phages is purified as described by Grossberger, Nucleic Acid Res. 15: 6737, 1987. A 5 kb EcoR I fragment, containing the whole CYP79B2 coding region and 2361 bp of the promoter region (see nucleotides 60536 to 62896 of GenBank Accession No. AL035708, SEQ ID NO: 16), is subcloned into pBluescript II SK (Stratagene). An *Xba* I restriction site is introduced by PCR immediately downstream of the CYP79B2 start codon using the T7 vector primer and the Xba I primer (Example 17). The PCR reaction contains 200 µM dNTPs, 400 pmol of each primer, 0.1 µg template DNA and 10 units *Pwo* polymerase in a total volume of 200 µl in *Pwo* polymerase PCR buffer with 2 mM $MgSO_4$ (Boehringer Mannheim).

primer 5 ..5'-AAAGCTCAATGCGTAGAAT-3' (SEQ ID NO: 7),

primer 6 ..5'-TTTTTAGACACCATCTTGTTTTCTTCTTC-3' (SEQ ID NO: 8),

primer 7 ..5'-TGTAGCGGCGCATTAAGC-3' (SEQ ID NO: 9),

primer 8 ..5'-CAAAAGAATAGACCGAGATAGGG-3' (SEQ ID NO: 10),

Example 24 - **CYP79F1 *E. coli* expression constructs**

CYP79F1 is one of several CYP79 homologues identified in the genome of *A. thaliana*. The deduced amino acid sequence of CYP79F1 has 88% identity with the deduced amino acid sequence of CYP79F2 and 43-50% identity with other CYP79 homologues from glucosinolate and cyanogenic glucoside containing species. CYP79F1 and CYP79F2 are located on the same chromosome, only separated by 1638 bp. This suggests that the two genes have been formed by gene duplication and might catalyze similar reactions. The expression construct is derived from the EST ATTS5112 (Arabidopsis Biological Resource Center, Ohio, USA) which contains the full length sequence of CYP79F1. The CYP79F1 coding region is amplified from the EST by PCR using primer 1 (sense direction) and primer 2 (antisense direction). Primer 1 introduces an *Xba*I site upstream of the start codon and an *Nde*I restriction site at the start codon. To optimize the construct for *E. coli* expression (Barnes et al, Proc. Natl. Acad. Sci. USA 88: 5597-5601, 1991) primer 1 changes the second codon from ATG to GCT and introduces a silent mutation in codon 5. Primer 2 introduces a *Bam*HI restriction site immediately after the stop codon. The PCR reaction is set up in a total volume of 50 µl in *Pwo* polymerase PCR buffer with 2 mM $MgSO_4$ using 2.5 units *Pwo* polymerase (Roche Molecular Biochemicals), 0.1 µg template DNA, 200 µM dNTPs and 50 pmol of each primer. After incubation of the reaction at 94°C for 5 min, 20 PCR cycles of 15 sec at 94°C, 30 sec at 58°C, and 2 min at 72°C are run. The PCR fragment is digested with *Xba*I and *Bam*HI, and ligated into the *Xba*I/*Bam*HI digested vector pBluescript II SK (Stratagene). The cDNA is sequenced on an ALF-Express (Pharmacia) using the Thermo Sequence Fluorescent-labelled Primer cycle sequencing kit (7-deaza dGTP) (Pharmacia) to exclude PCR errors and transferred from pBluescript II SK to an *Nde*I/*Bam*HI digested pSP19g10L expression vector (Barnes et al, Proc. Natl. Acad. Sci. USA 88: 5597-5601, 1991).

potential substrates. After incubation at 28°C for 1 hour, half of the reaction mixture is analyzed by TLC on Silica Gel 60 F_{254} sheets (Merck) using toluene/ethyl acetate 5:1 (v/v) as eluent. Radiolabelled bands are visualized and quantified using a STORM 840 phosphoimager (Pharmacia). For GC-MS analysis, 450 µl reaction mixture containing 3.3 mM L-methionine (Sigma), 3.3 mM DL-dihomomethionine or 3.3 mM DL-trihomomethionine, respectively, are incubated for 4 hours at 25°C and extracted with a total volume of 600 µl $CHCl_3$. The organic phase is collected, evaporated, and the residue is dissolved in 15 µl $CHCl_3$ and analyzed by GC-MS. GC-MS analysis is performed on an HP5890 Series II gas chromatograph directly coupled to a Jeol JMS-AX505W mass spectrometer. An SGE column (BPX5, 25 m x 0.25 mm, 0.25 µm film thickness) is used (heat pressure 100 kPa, splitless injection). The oven temperature program is as follows: 80°C for 3 minutes, 80°C to 180°C at 5°C min^{-1},180°C to 300°C at 20°C min^{-1}, and 300°C for 10 min. The ion source is run in EI mode (70 eV) at 200°C. The retention times of the *E*- and *Z*-isomer of 5-methylthiopentanaldoxime are 14.3 min and 14.8 min, respectively. The two isomers have identical fragmentation patterns with m/z values of 130, 129, 113, 82, 61 and 55 as the most prominent peaks. The retention times of the *E*- and *Z*-isomer of 6-methylthiopentanaldoxime are 17.1 min and 17.6 min, respectively. The two isomers have identical fragmentation patterns with m/z values of 144, 143, 98, 96, 69, 61 and 55 as the most prominent peaks. DL-dihomomethionine, DL-trihomomethionine, 5-methylthiopentanaldoxime and 6-methylthiohexanaldoxime are synthesized as described (Dawson et al, J. Biol. Chem. 268: 27154-27159, 1993) and authenticated by NMR spectroscopy.

A CO difference spectrum with the characteristic peak at 450 nm is obtained for CYP79F1 expressed in *E. coli* strain C43(DE3), but not for CYP79F1 expressed in *E. coli* strain JM109. In addition to the peak at 450 nm, a peak at 418 nm is detected.

To identify substrates of CYP79F1, activity measurements are carried out using spheroplasts of *E. coli* C43(DE3) reconstituted with NADPH:cytochrome P450 reductase from *S. bicolor*. When the reaction mixture containing CYP79F1 is incubated with DL-dihomomethionine, two compounds, which are not present in the control reactions, are detected by GC-MS. The retention times and the mass spectral fragmentation patterns of these compounds are identical with those for the E/Z-isomers of synthetic 5-methylthiopentanaldoxime. When DL-trihomomethionine is administred to the reaction mixture containing CYP79F1, two compounds with retention times and fragmentation

normal appearance within the first seven weeks of growth. Before floral transition becomes apparent, reduced apical dominance results in production of multiple axillary shoots which later developed into lateral inflorescences. These morphological changes give S5, S7 and S9 a bushy phenotype. In addition, S5 has curly rosette leaves with the leaf tips bending downwards.

Transgenic *A. thaliana* plants with altered content of aliphatic glucosinolates due to co-suppression or over-expression of CYP79F1 possess a characteristic morphological phenotype characterized by prolonged vegetative growth and production of multiple axillary shoots. *A. thaliana* has been reported to be able to tolerate overexpression of cytochromes P450 of the CYP79 family leading to a two to five fold increase in glucosinolate content without similar changes in the appearence of the plants. Therefore it seems unlikely that the morphological changes result from the presence or absense of specific glucosinolates. A possible explanation is that the morphological phenotype is due to a pleiotropic effect caused by disturbance of the plant's sulfur metabolism, in which methionine plays a central role. Alterations of the methionine metabolism may explain why both plants with co-suppression and overexpression of CYP79F1 show similar morphological changes when compared to wild-type plants. The onset of the morphological changes in CYP79F1 co-suppressed plants at the time of floral transition may be due to the requirement for methionine to support flower development. Alternatively, it coincides with an increase in the level of CYP79F1 expression in wild-type plants.

HPLC analysis of the glucosinolate content of plant extracts

Six to eight rosette leaves from each plant are harvested from nine 9-week-old primary transformants of 35S:CYP79F1 plants and ten 7-week-old wild-type plants of the same size. The tissue is immediately frozen in liquid nitrogen and freeze-dried for 48 hours. Glucosinolates are analyzed as desulfoglucosinolates as follows: 3.5 ml of boiling 70% (v/v) methanol are added to 9 to 20 mg freeze-dried material, 10 µL internal standard (5 mM *p*-hydroxybenzylglucosinolate; Bioraf, Denmark) are added, and the sample is incubated in a boiling water bath for 4 min. Plant material is pelleted, the pellet is re-extracted with 3.5 ml 70% (v/v) methanol and centrifuged. The supernatants are pooled and analyzed by HPLC after sulfatase treatment as described by Wittstock et al, J. Biol. Chem. 275, 14659-14666, 2000. The assignment of peaks is based on retention times and UV spectra compared to standard compounds. Glucosinolates are quantified in relation to the internal standard and by use of response factors (Haughn et al, Plant Physiol. 97: 217-226, 1991; Buchner in:

As the dihomomethionine-derived glucosinolates are the major glucosinolates of wild-type rosette leaves, altered levels of these glucosinolates influence the total glucosinolate content remarkably. This is particularly pronounced in the plants with CYP79F1 co-suppression. These plants have a total glucosinolate content ranging from 4.3 to 4.8 µmol $(g\ dw)^{-1}$ as compared to the total glucosinolate content of wild-type plants ranging from 8.8 to 17.4 µmol $(g\ dw)^{-1}$. In addition to the changes in the content of 4-methylsulfinylbutylglucosinolate and 4-methylthiobutyl-glucosinolate, alterations in the level of other glucosinolates, particularly of Methionine-derived glucosinolates, are observed in 35S:CYP79F1 plants. Plants with a reduced content of 4-methylsulfinylbutylglucosinolate and 4-methylthiobutylglucosinolate also have reduced levels of the other major glucosinolates derived from chain-elongated methionine homologues, i.e. 3-methylsulfinylpropylglucosinolate and 8-methylsulfinyloctylglucosinolate. This might be explained by co-suppression not only of the CYP79F1 transcript but also of transcripts of other CYP79 homologues involved in the biosynthesis of aliphatic glucosinolates such as transcripts of CYP79F2 which has 88% amino acid identity with CYP79F1. Alternatively, it might reflect that CYP79F1 has a broad substrate specificity for chain-elongated methionines. The fact that chain-elongated methionines accumulate in plants with CYP79F1 co-suppression indicates that the enzymes catalyzing the chain elongation of methionine are not subject to feedback inhibition by the chain-elongated product. The content of the three indoleglucosinolates is not affected significantly.

Analysis of the amino acid content of plant extracts

Rosette leaves from three 12-week-old primary transformants of 35S:CYP79F1 plants and three 8-week-old wild-type plants of the same size are used. 250 mg of leaf material from each plant are homogenized in 3 ml 50 mM KP_i, pH 7.5 using a Polytron homogenizer. The plant material is pelleted (20000g for 10 minutes) and re-extracted twice with 3 ml 50 mM KP_i, pH 7.5. The water phases are combined, dried *in vacuo*, and the residue is dissolved in 100 µl water. An aliquot of the redissolved extract is treated with 1/10 volume 30% salicylic sulfonic acid and denatured proteins are removed by centrifugation. The supernatant is neutralized with 1/10 volume 1 N NaOH. The individual protein amino acids in the sample are identified and quantified using an Ultropac 8 Resin Reverse Phase HPLC column (200 x 4.6 mm) on a Biochrom 20 amino acid analyzer (Pharmacia) essentially according to the manufacturer's elution program.

µM rNTPs, 10 mM DTT, 100 units RNAsin Ribonuclease inhibitor (Promega), 3 µg linearized pBluescript II SK, and 50 units T3 RNA polymerase (Promega). After incubation at 37°C for 2 hours, 20 units of RNase-free DNase are added, and the reaction is incubated at 37°C for another 1 hour. Following extraction with phenol and $CHCl_3$ and precipitation with ethanol, the RNA is dissolved in diethylpyrocarbonate-treated water.

The following tissues are harvested from *A. thaliana*:

(1) total plant tissue of 4-week-old plants (grown at 8 hours light/ 16 hours dark);
(2) rosette leaves (without petioles) and
(3) above ground parts of 5-week-old plants (before onset of floral transition; grown at 8 hours light/ 16 hours dark);
(4) rosette leaves (without petioles) and
(5) cauline leaves of flowering plants (9 weeks old; grown at 12 hours light/ 12 hours dark to induce flowering).

Total RNA is isolated from said tissuey using TRIZOL-Reagent (GIBCO BRL). The RNA is quantified spectrophotometrically and used to synthesize first-strand cDNA. To ensure linearity of the RT-PCR, first-strand cDNA synthesis is performed on 1 µg, 0.3 µg and 0.1 µg of each pool of RNA.The cDNA is synthesized in First Strand Buffer (GIBCO BRL) supplemented with 0.5 mM dNTPs, 10 mM DTT, 200 ng random hexamers (Pharmacia), 3 pg control RNA (internal standard), and 200 units SUPERSCRIPTII Reverse transcriptase (GIBCO BRL) in a total volume of 20 µl. The reaction mixture is incubated at 27°C for 10 minutes followed by incubation at 42°C for 50 minutes and inactivation at 95°C for 5 minutes. The RT-reactions are purified by means of a PCR-purification kit (QIAGEN; elution with 50 µl of 1 mM Tris-buffer, pH 8). 2 µl of the purified RT-reactions are subjected to PCR. The PCR reactions are set up in a total volume of 50 µl in PCR buffer (GIBCO BRL) supplemented with 200 µM dNTPs, 1.5 mM $MgCl_2$, 50 pmol of sense primer, 50 pmol of antisense primer, and 2.5 units Platinum *Taq* DNA polymerase (GIBCO BRL). The PCR program is as follows: 2 minutes at 94°C, 32 cycles of 30 seconds at 94°C, 30 seconds at 57°C, 50 seconds at 72°C. 10 µl of the PCR reactions are analyzed by gel electrophoresis on 1% agarose gels. Bands are visualized by ethidium bromide staining and quantified on a Gel Doc 2000 Transilluminator (Biorad). The primers used to analyze the CYP79F1 transcript are primer 5 (sense direction) and primer 6 (antisense direction). At 57°C primer 5 does not anneal to genomic DNA comprising the CYP79F1 gene as the sequence of primer 5 is complementary to the sequences flanking an 111 bp intron of the CYP79F1

What is claimed is:

1. A DNA coding for a P450 monooxygenase converting an aliphatic or aromatic amino acid or chain-elongated methionine homologue to the corresponding oxime.
2. The DNA of claim 1 converting L-Valine or L-Isoleucine to the corresponding oxime; tyrosine to *p*-hydroxyphenylacetaldoxime; L-phenylalanine to phenylacetaldoxime; tryptophan to indole-3-acetaldoxime; or chain-elongated methionine to the corresponding oxime.
3. The DNA of claim 1 coding for a P450 monooxygenase consisting of amino acid residues independently selected from the group of the amino acid residues Gly, Ala, Val, Leu, Ile, Phe, Pro, Ser, Thr, Cys, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg and His, wherein global alignment of the amino acid sequence of the encoded protein shows at least 40% identity to the amino acid sequence resulting from the global alignment with SEQ ID NO: 1 or SEQ ID NO: 3or both; SEQ ID NO: 39; or SEQ ID NO: 54 or SEQ ID NO: 70 or both; or at least 50% identity to the amino acid sequence resulting from the global alignment with SEQ ID NO: 9 or SEQ ID NO: 11 or both or SEQ ID NO: 74 or SEQ ID NO: 84 or both.
4. The DNA of claim 1, wherein an open reading frame is operably linked to one or more regulatory sequences different from the regulatory sequences associated with the genomic gene containing the exons of the open reading frame.
5. The DNA of claims 1 to 4 coding for a P450 monooxygenase having the formula R_1-R_2-R_3, wherein
   - R_1, R_2 and R_3 designate component sequences, and
   - R_2 consists of 150 to 175 or more amino acid residues the sequence of which is at least 60% to 65% identical to an aligned component sequence of SEQ ID NO: 1 or SEQ ID NO: 3; SEQ ID NO: 9 or SEQ ID NO: 11; SEQ ID NO: 39; SEQ ID NO: 54 or SEQ ID NO: 70; or SEQ ID NO: 74 or SEQ ID NO: 84.
6. The DNA of claim 1, wherein the amino acid sequence of R_2 is represented by amino acids 334-484 of SEQ ID NO: 1 or amino acids 333-483 of SEQ ID NO: 3; amino acids 339-489 of SEQ ID NO: 9 or amino acids 332-482 of SEQ ID NO: 11; amino acids 308-487 of SEQ ID NO: 39; amino acids 196-345 of SEQ ID NO: 54 or amino acids 192-341 of SEQ ID NO: 70; amino acids 334-483 of SEQ ID NO: 74 or amino acids 332-481 of SEQ ID NO: 84.

the amino acid sequence resulting from the global alignment with SEQ ID NO: 1 or SEQ ID NO: 3 or both; SEQ ID NO: 39; SEQ ID NO: 54 or SEQ ID NO: 70 or both; or at least 50% identity to the amino acid sequence resulting from the global alignment with SEQ ID NO: 9 or SEQ ID NO: 11 or both; or SEQ ID NO: 74 or SEQ ID NO: 84 or both;

(f) optionally further processing the purified DNA.

13. A marker assisted breeding method selecting plants with a desired trait using hybridization with one or more oligonucleotides, wherein the sequence of at least one of said oligonucleotides constitutes a component sequence of the DNA of claim 1.

14. A method for producing purified recombinant P450 monooxygenase converting an aliphatic or aromatic amino acid or chain-elongated methionine homologue to the corresponding oxime, comprising expression of a corresponding gene in *P. pastoris*.

15. A method for obtaining a transgenic plant , comprising

(a) stably integrating into a plant cell or tissue which can be regenerated to a complete plant DNA comprising at least part of an open reading frame of a P450 monooxygenase converting an aliphatic or aromatic amino acid or chain-elongated methionine homologue to the corresponding oxime , and

(b) selecting transgenic plants.

16. The method of claim 15 resulting in transgenic expression of a P450 monooxygenase in a plant.

17. The method of claim 15 resulting in the reduced expression of an endogenous P450 monooxygenase in a plant.

18. The method of claim 15 resulting in an altered content or profile of cyanogenic glucosides or glucosinolates.

SEQUENCE LISTING

<110> Novartis AG
Royal Veterinary and Agricultural University

<120> P450 monooxygenases of the CYP79 family

<130> S-31292A

<140>
<141>

<150> EP 00100646.9
<151> 2000-01-13

<150> EP 00107001.0
<151> 2000-03-30

<150> EP 00109423.4
<151> 2000-05-03

<150> EP 00114184.5
<151> 2000-07-13

<150> EP 00114912.9
<151> 2000-07-17

<160> 85

<170> PatentIn Ver. 2.1

<210> 1
<211> 542
<212> PRT
<213> Manihot esculenta

<400> 1

Met	Ala	Met	Asn	Val	Ser	Thr	Thr	Ile	Gly	Leu	Leu	Asn	Ala	Thr	Ser
1				5				10					15		
Phe	Ala	Ser	Ser	Ser	Ser	Ile	Asn	Thr	Val	Lys	Ile	Leu	Phe	Val	Thr
		20					25					30			
Leu	Phe	Ile	Ser	Ile	Val	Ser	Thr	Ile	Val	Lys	Leu	Gln	Lys	Ser	Ala
	35					40					45				
Ala	Asn	Lys	Glu	Gly	Ser	Lys	Lys	Leu	Pro	Leu	Pro	Pro	Gly	Pro	Thr
50					55					60					
Pro	Trp	Pro	Leu	Ile	Gly	Asn	Ile	Pro	Glu	Met	Ile	Arg	Tyr	Arg	Pro
65				70					75					80	
Thr	Phe	Arg	Trp	Ile	His	Gln	Leu	Met	Lys	Asp	Met	Asn	Thr	Asp	Ile
			85					90					95		
Cys	Leu	Ile	Arg	Phe	Gly	Arg	Thr	Asn	Phe	Val	Pro	Ile	Ser	Cys	Pro

100 105 110

Val Leu Ala Arg Glu Ile Leu Lys Lys Asn Asp Ala Ile Phe Ser Asn
115 120 125

Arg Pro Lys Thr Leu Ser Ala Lys Ser Met Ser Gly Gly Tyr Leu Thr
130 135 140

Thr Ile Val Val Pro Tyr Asn Asp Gln Trp Lys Lys Met Arg Lys Ile
145 150 155 160

Leu Thr Ser Glu Ile Ile Ser Pro Ala Arg His Lys Trp Leu His Asp
165 170 175

Lys Arg Ala Glu Glu Ala Asp Asn Leu Val Phe Tyr Ile His Asn Gln
180 185 190

Phe Lys Ala Asn Lys Asn Val Asn Leu Arg Thr Ala Thr Arg His Tyr
195 200 205

Gly Gly Asn Val Ile Arg Lys Met Val Phe Ser Lys Arg Tyr Phe Gly
210 215 220

Lys Gly Met Pro Asp Gly Gly Pro Gly Pro Glu Glu Ile Glu His Ile
225 230 235 240

Asp Ala Val Phe Thr Ala Leu Lys Tyr Leu Tyr Gly Phe Cys Ile Ser
245 250 255

Asp Phe Leu Pro Phe Leu Leu Gly Leu Asp Leu Asp Gly Gln Glu Lys
260 265 270

Phe Val Leu Asp Ala Asn Lys Thr Ile Arg Asp Tyr Gln Asn Pro Leu
275 280 285

Ile Asp Glu Arg Ile Gln Gln Trp Lys Ser Gly Glu Arg Lys Glu Met
290 295 300

Glu Asp Leu Leu Asp Val Phe Ile Thr Leu Lys Asp Ser Asp Gly Asn
305 310 315 320

Pro Leu Leu Thr Pro Asp Glu Ile Lys Asn Gln Ile Ala Glu Ile Met
325 330 335

Ile Ala Thr Val Asp Asn Pro Ser Asn Ala Ile Glu Trp Ala Met Gly
340 345 350

Glu Met Leu Asn Gln Pro Glu Ile Leu Lys Lys Ala Thr Glu Glu Leu
355 360 365

Asp Arg Val Val Gly Lys Asp Arg Leu Val Gln Glu Ser Asp Ile Pro
370 375 380

Asn Leu Asp Tyr Val Lys Ala Cys Ala Arg Glu Ala Phe Arg Leu His
385 390 395 400

Pro Val Ala His Phe Asn Val Pro His Val Ala Met Glu Asp Thr Val
405 410 415

Ile Gly Asp Tyr Phe Ile Pro Lys Gly Ser Trp Ala Val Leu Ser Arg
420 425 430

Tyr Gly Leu Gly Arg Asn Pro Lys Thr Trp Ser Asp Pro Leu Lys Tyr
435 440 445

Asp Pro Glu Arg His Met Asn Glu Gly Glu Val Val Leu Thr Glu His
450 455 460

Glu Leu Arg Phe Val Thr Phe Ser Thr Gly Arg Arg Gly Cys Val Ala
465 470 475 480

Ser Leu Leu Gly Ser Cys Met Thr Thr Met Leu Leu Ala Arg Met Leu
485 490 495

Gln Cys Phe Thr Trp Thr Pro Pro Ala Asn Val Ser Lys Ile Asp Leu
500 505 510

Ala Glu Thr Leu Asp Glu Leu Thr Pro Ala Thr Pro Ile Ser Ala Phe
515 520 525

Ala Lys Pro Arg Leu Ala Pro His Leu Tyr Pro Thr Ser Pro
530 535 540

<210> 2
<211> 1845
<212> DNA
<213> Manihot esculenta

<400> 2
gttcagggca tatcaatatg gccatgaacg tctccaccac catcggttta cttaacgcca 60
cctccttcgc ctcctcctcc tccatcaaca cggtcaagat cttgttcgtc accctcttta 120
tttccattgt tagtactatt gtaaaacttc aaaagagtgc tgctaacaag gaaggtagca 180
agaaactccc actccctcct ggccctactc catggccact catcggaaac atcccggaaa 240
tgatccggta cagacccacg tttcggtgga ttcaccaact catgaaggac atgaacactg 300
atatttgtct cattcgtttt ggaagaacta actttgttcc tataagctgt cctgttcttg 360
ctcgtgaaat actaaaaaag aatgacgcta tcttctctaa caggccaaag actctctctg 420
caaaatctat gagcggagga tacttgacaa ctattgtggt gccatacaat gaccaatgga 480
agaaaatgag gaagatctta acctcagaga tcatttctcc ggccagacac aaatggctcc 540
atgataaaag agctgaggag gctgataatc ttgtgttcta catccacaac cagttcaaag 600
caaataaaaa tgtgaatttg agaacagcca ccaggcatta cggcgggaat gtgatcagaa 660
aaatggtgtt cagcaagaga tacttcggca agggaatgcc ggacggagga ccagggcctg 720
aagaaatcga gcacattgat gccgttttca ctgccttgaa atacttgtat gggttttgca 780
tatcagattt cttgcctttc ttgttgggac ttgatctgga tggccaagaa aaatttgtgc 840
ttgatgcaaa taagaccata agggattatc agaacccttt aattgatgaa aggattcaac 900
aatggaagag tggtgaaagg aaggaaatgg aggacttgct tgatgttttc atcactctca 960
aggattcaga cggcaaccca ttgctcactc ctgacgagat caagaatcaa atagctgaaa 1020
ttatgatagc aacagtagat aacccatcaa acgcaatcga atgggcaatg ggggagatgc 1080
taaatcaacc agaaatcctg aagaaggcca cagaagagct cgacagggtg gtcggcaaag 1140
acaggcttgt tcaagaatcc gacatcccca accttgacta tgtcaaagcc tgtgcaagag 1200
aagccttcag gctccatcca gtagcacact tcaatgtccc tcatgtagcc atggaagaca 1260
ctgtcattgg tgattacttt attccaaagg gcagctgggc agttctcagc cgctatgggc 1320

tcggcaggaa cccaaagaca tggtctgatc ctctcaagta cgatccagaa aggcacatga 1380
acgagggaga ggtggtgctc actgagcacg agttaaggtt tgtgactttc agcactggaa 1440
gacgtggctg cgtagcttcg ttgcttggaa gctgcatgac gacgatgttg ctggcgagga 1500
tgctgcagtg cttcacttgg actccaccag ccaatgtttc caagattgat ctcgccgaga 1560
ctctagatga gcttactcct gcaacaccca tctctgcatt tgccaagcct cgcctggctc 1620
ctcatctcta cccaacgtca ccttgaaaga gagatcagat cttatcagtt cttagaacgt 1680
cctttaatta tgatttgcta aaaacaaata aaaatcattt ggttattgtg taggtaatct 1740
tacaagcttc ctgtttattg agagttgtta attaactctc aaaatgattt gtggggttat 1800
cttgtttctc ttgcaatata gttgctttac tagaaaaaaa aaaaa 1845

<210> 3
<211> 541
<212> PRT
<213> Manihot esculenta

<400> 3

```
Met Ala Met Asn Val Ser Thr Thr Ala Thr Thr Thr Ala Ser Phe Ala
  1               5                  10                  15

Ser Thr Ser Ser Met Asn Asn Thr Ala Lys Ile Leu Leu Ile Thr Leu
             20                  25                  30

Phe Ile Ser Ile Val Ser Thr Val Ile Lys Leu Gln Lys Arg Ala Ser
         35                  40                  45

Tyr Lys Lys Ala Ser Lys Asn Phe Pro Leu Pro Pro Gly Pro Thr Pro
     50                  55                  60

Trp Pro Leu Ile Gly Asn Ile Pro Glu Met Ile Arg Tyr Arg Pro Thr
 65                  70                  75                  80

Phe Arg Trp Ile His Gln Leu Met Lys Asp Met Asn Thr Asp Ile Cys
                 85                  90                  95

Leu Ile Arg Phe Gly Lys Thr Asn Val Val Pro Ile Ser Cys Pro Val
            100                 105                 110

Ile Ala Arg Glu Ile Leu Lys Lys His Asp Ala Val Phe Ser Asn Arg
        115                 120                 125

Pro Lys Ile Leu Cys Ala Lys Thr Met Ser Gly Gly Tyr Leu Thr Thr
    130                 135                 140

Ile Val Val Pro Tyr Asn Asp Gln Trp Lys Lys Met Arg Lys Val Leu
145                 150                 155                 160

Thr Ser Glu Ile Ile Ser Pro Ala Arg His Lys Trp Leu His Asp Lys
                165                 170                 175

Arg Ala Glu Glu Ala Asp Gln Leu Val Phe Tyr Ile Asn Asn Gln Tyr
            180                 185                 190

Lys Ser Asn Lys Asn Val Asn Val Arg Ile Ala Ala Arg His Tyr Gly
        195                 200                 205
```

Gly Asn Val Ile Arg Lys Met Met Phe Ser Lys Arg Tyr Phe Gly Lys
210 215 220

Gly Met Pro Asp Gly Gly Pro Gly Pro Glu Glu Ile Met His Val Asp
225 230 235 240

Ala Ile Phe Thr Ala Leu Lys Tyr Leu Tyr Gly Phe Cys Ile Ser Asp
245 250 255

Tyr Leu Pro Phe Leu Glu Gly Leu Asp Leu Asp Gly Gln Glu Lys Ile
260 265 270

Val Leu Asn Ala Asn Lys Thr Ile Arg Asp Leu Gln Asn Pro Leu Ile
275 280 285

Glu Glu Arg Ile Gln Gln Trp Arg Ser Gly Glu Arg Lys Glu Met Glu
290 295 300

Asp Leu Leu Asp Val Phe Ile Thr Leu Gln Asp Ser Asp Gly Lys Pro
305 310 315 320

Leu Leu Asn Pro Asp Glu Ile Lys Asn Gln Ile Ala Glu Ile Met Ile
325 330 335

Ala Thr Ile Asp Asn Pro Ala Asn Ala Val Glu Trp Ala Met Gly Glu
340 345 350

Leu Ile Asn Gln Pro Glu Leu Leu Ala Lys Ala Thr Glu Glu Leu Asp
355 360 365

Arg Val Val Gly Lys Asp Arg Leu Val Gln Glu Ser Asp Ile Pro Asn
370 375 380

Leu Asn Tyr Val Lys Ala Cys Ala Arg Glu Ala Phe Arg Leu His Pro
385 390 395 400

Val Ala Tyr Phe Asn Val Pro His Val Ala Met Glu Asp Ala Val Ile
405 410 415

Gly Asp Tyr Phe Ile Pro Lys Gly Ser Trp Ala Ile Leu Ser Arg Tyr
420 425 430

Gly Leu Gly Arg Asn Pro Lys Thr Trp Pro Asp Pro Leu Lys Tyr Asp
435 440 445

Pro Glu Arg His Leu Asn Glu Gly Glu Val Val Leu Thr Glu His Asp
450 455 460

Leu Arg Phe Val Thr Phe Ser Thr Gly Arg Arg Gly Cys Val Ala Ala
465 470 475 480

Leu Leu Gly Thr Thr Met Ile Thr Met Met Leu Ala Arg Met Leu Gln
485 490 495

Cys Phe Thr Trp Thr Pro Pro Pro Asn Val Thr Arg Ile Asp Leu Ser
500 505 510

Glu Asn Ile Asp Glu Leu Thr Pro Ala Thr Pro Ile Thr Gly Phe Ala
515 520 525

Lys Pro Arg Leu Ala Pro His Leu Tyr Pro Thr Ser Pro
530 535 540

<210> 4
<211> 1920
<212> DNA
<213> Manihot esculenta

<400> 4
ggtcttggtc atagccctgg acttgaattg ttcagggcaa caccaatatg gccatgaacg 60
tctccaccac cgcaaccacc acggcctcct tcgcctccac gtcctccatg aacaatactg 120
ccaaaatcct ccttatcacc ctcttcattt ccattgtcag tactgttata aaacttcaaa 180
aaagggcatc ctacaagaaa gctagcaaga acttcccact ccctcctggt ccgactccat 240
ggccactcat cggaaacatc cctgaaatga tccggtacag accgacgttt cgttggattc 300
accaactcat gaaggacatg aacaccgata tttgtctgat ccgtttcgga aaaactaacg 360
ttgttcctat tagctgccct gtcattgctc gtgaaatcct gaaaaagcac gatgctgtct 420
tctctaacag gccaaagatt ctctgcgcta aaacaatgag cggcggatac ttgacgacga 480
ttgtggtgcc atacaatgat caatggaaga aaatgaggaa ggtcctaact tcagagatca 540
tttctccagc taggcacaaa tggctccatg ataagagagc tgaggaagca gatcagcttg 600
tgttctatat caataaccag tacaagagca acaagaatgt gaatgtgaga attgcggcaa 660
ggcattacgg tggaaatgtg atcagaaaga tgatgtttag caagagatac ttcggcaaag 720
ggatgcctga tggaggacca gggcctgaag aaatcatgca cgttgatgca atttttacag 780
cacttaaata tttgtatgga ttttgcatct ctgattactt gccttttttg gaggggcttg 840
atcttgatgg ccaggaaaag attgtgctta atgcaaataa gaccataagg gatcttcaaa 900
acccattaat agaagaaagg attcaacaat ggaggagtgg tgaaagaaag gaaatggaag 960
acttgcttga tgttttcatt actcttcagg attcagatgg caagccattg ctcaatccag 1020
acgagataaa gaatcaaatc gctgaaatta tgatagcaac aatagacaac ccagcaaacg 1080
ccgtagaatg ggcaatgggg gagctgataa atcaaccaga acttctggca aaggccacag 1140
aggaacttga cagagtggtc ggcaaagaca ggcttgtgca agaatctgac atccctaatc 1200
ttaattacgt caaagcctgt gcaagggagg ccttcaggct ccacccagtt gcatacttca 1260
acgtccctca cgtagccatg gaagacgccg tcatcggcga ttacttcatt ccaaagggca 1320
gctgggcaat tcttagccgc tacgggctcg gccggaaccc aaaaacatgg cctgatccac 1380
tcaagtacga cccagaaagg cacttgaacg agggcgaagt ggtgctgact gagcacgacc 1440
ttaggttcgt cacattcagc actggacgtc gtgggtgtgt cgctgctttg cttggaacca 1500
ccatgattac gatgatgctg gccaggatgc ttcagtgctt cacttggact ccacccccta 1560
atgtaaccag gattgatctc agtgagaata tcgatgagct tactccagca acacccatca 1620
ctggatttgc taagccacgg ttggctcctc atctctaccc cacttcacct tgaattaaag 1680
cccaaagatg ggaagggatg aatgtgagtt gttagaagtt ttaataaaaa aattattggg 1740
tttatatgtg taattacgtg gtaaccttac aaagtgtctg ttattgagag ttttaatctc 1800
tcaaaataat ttgtgtggct aagatttctt catctttgta tctcttgcaa ttgtttgctc 1860
tataaaacat cttatttcct taaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1920

<210> 5
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<221> modified_base
<222> (14)

<223> i

<220>
<221> modified_base
<222> (20)
<223> i

<220>
<221> modified_base
<222> (23)
<223> i

<220>
<223> Description of Artificial Sequence:
Oligonucleotide sequence

<400> 5
gcggaattca rggnaayccn ytnct 25

<210> 6
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<221> modified_base
<222> (18)
<223> i

<220>
<223> Description of Artificial Sequence:
Oligonucleotide sequence

<400> 6
cgcggatccg gdatrtcnga ytcytg 26

<210> 7
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide sequence

<400> 7
cgaaacgatg gctatgaacg tctct 25

<210> 8
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
Oligonucleotide sequence

<400> 8
tggtagagac gttcatagcc atcgttt 27

<210> 9
<211> 540
<212> PRT
<213> Triglochin maritima

<400> 9

```
Met Glu Leu Ile Thr Ile Leu Pro Ser Val Leu Pro Asn Ile His Ser
 1               5                  10                  15

Thr Ala Thr Val Leu Phe Leu Leu Leu Leu Thr Thr Ala Leu Ser Phe
            20                  25                  30

Leu Phe Leu Phe Lys Gln His Leu Thr Lys Leu Thr Lys Ser Lys Ser
        35                  40                  45

Lys Ser Thr Thr Leu Pro Pro Gly Pro Arg Pro Trp Pro Ile Val Gly
    50                  55                  60

Ser Leu Val Ser Met Tyr Met Asn Arg Pro Ser Phe Arg Trp Ile Leu
65                  70                  75                  80

Ala Gln Met Glu Gly Arg Arg Ile Gly Cys Ile Arg Leu Gly Gly Val
                85                  90                  95

His Val Val Pro Val Asn Cys Pro Glu Ile Ala Arg Glu Phe Leu Lys
            100                 105                 110

Val His Asp Ala Asp Phe Ala Ser Arg Pro Val Thr Val Val Thr Arg
        115                 120                 125

Tyr Ser Ser Arg Gly Phe Arg Ser Ile Ala Val Val Pro Leu Gly Glu
    130                 135                 140

Gln Trp Lys Lys Met Arg Arg Val Val Ala Ser Glu Ile Ile Asn Ala
145                 150                 155                 160

Lys Arg Leu Gln Trp Gln Leu Gly Leu Arg Thr Glu Glu Ala Asp Asn
                165                 170                 175

Ile Met Arg Tyr Ile Thr Tyr Gln Cys Asn Thr Ser Gly Asp Thr Asn
            180                 185                 190

Gly Ala Ile Ile Asp Val Arg Phe Ala Leu Arg His Tyr Cys Ala Asn
        195                 200                 205

Val Ile Arg Arg Met Leu Phe Gly Lys Arg Tyr Phe Gly Ser Gly Gly
    210                 215                 220
```

Glu Gly Gly Gly Pro Gly Lys Glu Glu Ile Glu His Val Asp Ala Thr
225 230 235 240

Phe Asp Val Leu Gly Leu Ile Tyr Ala Phe Asn Ala Ala Asp Tyr Val
245 250 255

Ser Trp Leu Lys Phe Leu Asp Leu His Gly Gln Glu Lys Lys Val Lys
260 265 270

Lys Ala Ile Asp Val Val Asn Lys Tyr His Asp Ser Val Ile Glu Ser
275 280 285

Arg Arg Glu Arg Lys Val Glu Gly Arg Glu Asp Lys Asp Pro Glu Asp
290 295 300

Leu Leu Asp Val Leu Leu Ser Leu Lys Asp Ser Asn Gly Lys Pro Leu
305 310 315 320

Leu Asp Val Glu Glu Ile Lys Ala Gln Ile Ala Asp Leu Thr Tyr Ala
325 330 335

Thr Val Asp Asn Pro Ser Asn Ala Val Glu Trp Ala Leu Ala Glu Met
340 345 350

Leu Asn Asn Pro Asp Ile Leu Gln Lys Ala Thr Asp Glu Val Asp Gln
355 360 365

Val Val Gly Arg His Arg Leu Val Gln Glu Ser Asp Phe Pro Asn Leu
370 375 380

Pro Tyr Ile Arg Ala Cys Ala Arg Glu Ala Leu Arg Leu His Pro Val
385 390 395 400

Ala Ala Phe Asn Leu Pro His Val Ser Leu Arg Asp Thr His Val Ala
405 410 415

Gly Phe Phe Ile Pro Lys Gly Ser His Val Leu Leu Ser Arg Val Gly
420 425 430

Leu Gly Arg Asn Pro Lys Val Trp Asp Asn Pro Leu Arg Phe Asp Pro
435 440 445

Asp Arg His Leu His Gly Gly Pro Thr Ala Lys Val Glu Leu Ala Glu
450 455 460

Pro Glu Leu Arg Phe Val Ser Phe Thr Thr Gly Arg Arg Gly Cys Met
465 470 475 480

Gly Gly Pro Leu Gly Thr Ala Met Thr Tyr Met Leu Leu Ala Arg Phe
485 490 495

Val Gln Gly Phe Thr Trp Gly Leu Arg Pro Ala Val Glu Lys Val Glu
500 505 510

Leu Glu Glu Glu Lys Cys Ser Met Phe Leu Gly Lys Pro Leu Arg Ala
515 520 525

Leu Ala Lys Pro Arg Gln Glu Leu Leu Gln Ser Phe
530 535 540

<210> 10
<211> 1858
<212> DNA
<213> Triglochin maritima

<400> 10
caatgcattg ctcccactag cccactacgt actataaatg catgcaccac tccacctctc 60
ctcctcagta gcaaaatgga actcataacc attcttccat cagtgcttcc taacatccac 120
tctactgcca cagtactgtt cctcttgcta ctcaccacag ccctctcctt cctcttcctc 180
ttcaaacaac acctcactaa gctaaccaag tccaagtcca agtccaccac attgccaccc 240
ggcccccgac catggcccat cgttggcagc ctcgtgtcga tgtacatgaa ccggccgtct 300
ttccggtgga tactagccca gatggagggg agaaggatag ggtgcattag gttgggtggt 360
gttcatgttg ttccggttaa ttgtcctgag attgctaggg agtttcttaa ggtgcatgat 420
gctgattttg catcgcgtcc ggtcacggtt gtgactcgct actcgtctcg tgggttccgg 480
tctattgccg tggttccact gggggagcaa tggaagaaga tgaggagggt ggtggcgtcg 540
gagattatta atgctaagag gctccaatgg cagcttgggc ttagaaccga agaagccgac 600
aacataatga ggtacatcac ctaccaatgc aacacttcgg gcgacactaa cggagcgatt 660
atcgacgtcc gcttcgccct ccgccactac tgtgccaatg tcatccggcg aatgctgttc 720
gggaaacgct acttcggaag cggtggagaa ggcggtgggc cgggaaagga ggagattgag 780
cacgttgacg ccaccttcga cgtcttgggt ctaatatacg ccttcaatgc ggcggactac 840
gtgtcgtggt tgaagttctt agacttgcat gggcaggaga agaaggttaa gaaggccatt 900
gatgtggtga ataagtatca tgactccgtt atcgagtcga ggagggagag gaaagtagag 960
ggaagagagg acaaggatcc agaggatctt cttgatgtgc ttttgtcgct taaggattct 1020
aatgggaagc ctctcttgga cgtggaggag atcaaagcac aaattgcgga tttgacgtac 1080
gcaacagttg ataacccgtc gaacgccgtg gaatgggcac tagccgagat gctgaacaac 1140
ccggacatcc tccaaaaggc gaccgacgag gtagaccagg tcgtcggaag gcaccgtctc 1200
gtacaagaat ccgacttccc gaacctcccc tacatccggg cctgcgcccg ggaggccctc 1260
cgtctccacc ctgtcgcggc cttcaacctc ccccacgtgt cccttcgtga cactcatgtc 1320
gccggttttt tcattccaaa aggcagccac gttctcctga gtcgcgtcgg cctcggacgc 1380
aaccccaagg tctgggacaa cccgcttcga ttcgaccccg accgacacct ccacggcggg 1440
cccaccgcca aagtcgagct ggccgagccg gagctgaggt tcgtgtcgtt caccaccggg 1500
aggagagggt gcatgggggg cccacttggg actgccatga cttatatgct gcttgctagg 1560
ttcgtccagg gtttcacttg gggtcttcgc cctgctgtgg agaaggttga gcttgaggag 1620
gagaagtgta gcatgttctt gggcaagcca ttaagggctt tggctaagcc acgtcaggag 1680
ctgctccaga gcttctaatt agggttaggg tttgggttgg attaataata cttatgaaat 1740
gcacgtttat gagtctataa atattatcca tgtaagtgtt atatgttttc gtgcaatcct 1800
attatccatg taagttaaat ttgataccat gaatgagttt atatgtgaaa aaaaaaaa 1858

<210> 11
<211> 533
<212> PRT
<213> Triglochin maritima

<400> 11
Leu Ile Thr Ile Leu Pro Ser Val Leu Pro Asn Ile His Ser Ser Ala
1 5 10 15

Thr Leu Phe Leu Leu Leu Leu Met Thr Thr Ala Leu Ser Phe Leu Phe
20 25 30

Leu Phe Lys Gln His Leu Ala Lys Leu Thr Lys Pro Lys Ser Thr Thr
35 40 45

Leu Pro Pro Gly Pro Arg Pro Trp Pro Ile Val Gly Ser Leu Val Ser
50 55 60

Met Tyr Met Asn Arg Pro Ser Phe Arg Trp Ile Leu Ala Gln Met Glu
65 70 75 80

Gly Arg Arg Ile Gly Cys Ile Arg Leu Gly Gly Val His Val Val Pro
85 90 95

Val Asn Cys Pro Glu Ile Ala Arg Glu Phe Leu Lys Val His Asp Ser
100 105 110

Asp Phe Ala Ser Arg Pro Val Thr Val Val Thr Arg Tyr Ser Ser Arg
115 120 125

Gly Phe Arg Ser Ile Ala Val Val Pro Leu Gly Glu Gln Trp Lys Lys
130 135 140

Met Arg Arg Val Val Ala Ser Glu Ile Ile Asn Ala Lys Arg Leu Gln
145 150 155 160

Trp Gln Leu Gly Leu Arg Thr Glu Glu Ala Asp Asn Ile Val Arg Tyr
165 170 175

Ile Thr Tyr Gln Cys Asn Thr Ser Gly Asp Thr Ser Gly Ala Ile Ile
180 185 190

Asp Val Arg Phe Ala Leu Arg His Tyr Cys Ala Asn Val Ile Arg Arg
195 200 205

Met Leu Phe Gly Lys Arg Tyr Phe Gly Ser Gly Gly Val Gly Gly Gly
210 215 220

Pro Gly Lys Glu Glu Ile Glu His Val Asp Ala Thr Phe Asp Val Leu
225 230 235 240

Gly Leu Ile Tyr Ala Phe Asn Ala Ala Asp Tyr Val Ser Trp Leu Lys
245 250 255

Phe Leu Asp Leu His Gly Gln Glu Lys Lys Val Lys Lys Ala Ile Asp
260 265 270

Val Val Asn Lys Tyr His Asp Ser Val Ile Asp Ala Arg Thr Glu Arg
275 280 285

Lys Val Glu Asp Lys Asp Pro Glu Asp Leu Leu Asp Val Leu Phe Ser
290 295 300

Leu Lys Asp Ser Asn Gly Lys Pro Leu Leu Asp Val Glu Glu Ile Lys
305 310 315 320

Ala Gln Ile Ala Asp Leu Thr Tyr Ala Thr Val Asp Asn Pro Ser Asn
325 330 335

```
Ala Val Glu Trp Ala Leu Ala Glu Met Leu Asn Asn Pro Ala Ile Leu
            340                 345                 350

Gln Lys Ala Thr Asp Glu Leu Asp Gln Val Val Gly Arg His Arg Leu
        355                 360                 365

Val Gln Glu Ser Asp Phe Pro Asn Leu Pro Tyr Ile Arg Ala Cys Ala
    370                 375                 380

Arg Glu Ala Leu Arg Leu His Pro Val Ala Ala Phe Asn Leu Pro His
385                 390                 395                 400

Val Ser Leu Arg Asp Thr His Val Ala Gly Phe Phe Ile Pro Lys Gly
                405                 410                 415

Ser His Val Leu Leu Ser Arg Val Gly Leu Gly Arg Asn Pro Lys Val
            420                 425                 430

Trp Asp Asn Pro Leu Gln Phe Asn Pro Asp Arg His Leu His Gly Gly
        435                 440                 445

Pro Thr Ala Lys Val Glu Leu Ala Glu Pro Glu Leu Arg Phe Val Ser
    450                 455                 460

Phe Thr Thr Gly Arg Arg Gly Cys Met Gly Gly Leu Leu Gly Thr Ala
465                 470                 475                 480

Met Thr Tyr Met Leu Leu Ala Arg Phe Val Gln Gly Phe Thr Trp Gly
                485                 490                 495

Leu His Pro Ala Val Glu Lys Val Glu Leu Gln Glu Glu Lys Cys Ser
            500                 505                 510

Met Phe Leu Gly Glu Pro Leu Arg Ala Phe Ala Lys Pro Arg Leu Glu
        515                 520                 525

Leu Leu Gln Ser Phe
    530
```

<210> 12
<211> 1778
<212> DNA
<213> Triglochin maritima

<400> 12
ctcataacca ttcttccatc agtgctacca aacatccact cttctgccac attgttcctc 60
ttgctactca tgaccacagc cctctccttc ctcttcctct tcaaacaaca cctcgctaag 120
ctaaccaaac ccaagtccac cacattgcca cctggccccc gaccctggcc catcgttggc 180
agcctcgtgt cgatgtacat gaaccggccg tccttccggt ggatactagc ccagatggag 240
gggaggagga tagggtgcat taggttgggt ggtgttcatg ttgttccggt taattgtcct 300
gagattgcta gggagtttct taaggtgcat gattctgatt ttgcatcgcg tccggtcacg 360
gttgtgactc gctactcgtc tcgtgggttc cggtctattg ccgtggttcc actgggggag 420
cagtggaaga agatgaggag ggtggtggca tcggagatta ttaatgctaa gaggctccaa 480
tggcagcttg ggcttagaac cgaagaagcc gacaacatag tgaggtacat cacctaccaa 540

tgcaacactt cgggcgacac tagcggagcg attatcgacg tccgcttcgc cctccgccac 600
tactgtgcca atgtcatccg gcgaatgctg ttcggaaaac gctactttgg tagcggtgga 660
gtaggcggtg ggcctggaaa ggaggagatt gagcacgttg acgccacctt cgacgtcttg 720
ggtctaatat acgccttcaa tgcggcggac tacgtgtcgt ggttgaagtt cttagacttg 780
catgggcagg agaagaaggt taagaaggcc attgatgtgg tgaataagta tcatgactcc 840
gttatcgacg cgaggacaga gagaaaagtg gaggataagg atccagagga tcttcttgat 900
gtgcttttt cgcttaagga ttctaatgga aagcctctct tggacgtgga ggagatcaaa 960
gcacaaattg cggatttgac gtacgcaaca gttgacaacc cgtcgaacgc cgtggaatgg 1020
gcactagccg agatgctgaa caacccggcc atcctccaaa aggcgaccga cgagctagac 1080
caggtcgtcg gaaggcaccg tctcgtacaa gaatccgact tcccgaacct cccctacatc 1140
cgtgcctgcg cccgggaggc cctccgtctc cacccggtcg cggctttcaa cctcccccac 1200
gtgtcccttc gtgacactca cgtcgccggc ttctttattc ccaaaggcag ccacgttctc 1260
ctgagtcgcg ttggcctcgg acgcaacccc aaggtgtggg acaacccgct tcaattcaac 1320
ccagaccgac acctccacgg cgggcccacc gccaaagtcg agctggccga accggagctg 1380
aggttcgtgt cgttcaccac cgggaggaga gggtgcatgg ggggcctact tgggactgcc 1440
atgacttata tgctgcttgc taggttcgtc cagggtttca cttgggggct tcaccctgct 1500
gtggagaagg ttgagcttca ggaggagaag tgtagcatgt tcttgggcga gccattgaga 1560
gcttttgcta agccacgtct ggagctgctc cagagcttct aattagtttt ggattaataa 1620
taactataat tactaccgat gtccttaaag ttgcatgtcg tgtaactagc acttgttata 1680
tttatagtta tgaaaggtac gtttatgaat ctataaaaat tatccatgta attgttatat 1740
gttttcgtgc aatcgtattg tgagtttggt ttacaaaa 1778

<210> 13
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<220>
<221> modified_base
<222> (18)
<223> i

<220>
<221> modified_base
<222> (21)
<223> i

<400> 13
gcggaattcg ayaayccnws naaygc 26

<210> 14
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<220>
<221> modified_base
<222> (11)

<223> i

<220>
<221> modified_base
<222> (17)
<223> i

<220>
<221> modified_base
<222> (20)
<223> i

<400> 14
gcggatccgc nacrtgnggn ahrttraa 28

<210> 15
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<220>
<221> modified_base
<222> (12)
<223> i

<220>
<221> modified_base
<222> (18)
<223> i

<220>
<221> modified_base
<222> (21)
<223> i

<400> 15
gcggaattcw snaaygcnrt ngartgg 27

<210> 16
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<220>
<221> modified_base
<222> (15)..(17)
<223> i

<220>
<221> modified_base
<222> (21)
<223> i

<220>
<221> modified_base
<222> (24)
<223> i

<400> 16
gcggatccrt traannnngc nacnggrtg 29

<210> 17
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 17
gcggaattcc acacaggaaa cagctatgac 30

<210> 18
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 18
gcggatccag acgagtagcg agtcacaac 29

<210> 19
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 19
gcggatccaa gaggaacagt act 23

<210> 20
<211> 23
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: primer

<400> 20
gcggatccaa gaggaacaat gtg 23

<210> 21
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 21
gcgaatgcat tgctcccact agcc 24

<210> 22
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 22
gcgatggtta tgagttccat tttg 24

<210> 23
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 23
gcgcatatgg aactaataac aattctt 27

<210> 24
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 24
gcgaagctta ttagaagctc tggagcag 28

<210> 25
<211> 51

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 25
gcgcatatgg ctctgttatt agcagttttt ttcctcttcc tcttcaaaca a 51

<210> 26
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 26
gcgcatatgg ctcgtcaagt tcattcttct tggaatttac caccaggccc c 51

<210> 27
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer encoded

<400> 27
Asp Asn Pro Ser Asn Ala
1 5

<210> 28
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer encoded

<220>
<221> VARIANT
<222> (3)
<223> V or L

<400> 28
Phe Asn Xaa Pro His Val Ala
1 5

<210> 29
<211> 6
<212> PRT

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer encoded

<400> 29
Ser Asn Ala Val Glu Trp
1 5

<210> 30
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer encoded

<400> 30
His Pro Val Ala Xaa Phe Asn
1 5

<210> 31
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer encoded

<400> 31
Val Val Thr Arg Tyr Ser Ser
1 5

<210> 32
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer encoded

<400> 32
Thr Val Leu Phe Leu Leu
1 5

<210> 33
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer encoded

<400> 33
Ala Thr Leu Phe Leu Leu
1 5

<210> 34
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer encoded

<400> 34
Met Glu Leu Ile Thr Ile
1 5

<210> 35
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer encoded

<400> 35
Met Glu Leu Ile Thr Ile Leu
1 5

<210> 36
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer encoded

<400> 36
Leu Leu Gln Ser Phe
1 5

<210> 37
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer encoded

<400> 37
Met Ala Leu Leu Leu Ala Val Phe Phe Leu Phe Leu Phe Lys Gln
1 5 10 15

<210> 38
<211> 15
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer encoded

<400> 38
Met Ala Arg Gln Val His Ser Ser Trp Asn Leu Pro Pro Gly Pro
1 5 10 15

<210> 39
<211> 523
<212> PRT
<213> Arabidopsis thaliana

<400> 39
Met Leu Ala Phe Ile Ile Gly Leu Leu Leu Leu Ala Leu Thr Met Lys
1 5 10 15

Arg Lys Glu Lys Lys Lys Thr Met Leu Ile Ser Pro Thr Arg Asn Leu
20 25 30

Ser Leu Pro Pro Gly Pro Lys Ser Trp Pro Leu Ile Gly Asn Leu Pro
35 40 45

Glu Ile Leu Gly Arg Asn Lys Pro Val Phe Arg Trp Ile His Ser Leu
50 55 60

Met Lys Glu Leu Asn Thr Asp Ile Ala Cys Ile Arg Leu Ala Asn Thr
65 70 75 80

His Val Ile Pro Val Thr Ser Pro Arg Ile Ala Arg Glu Ile Leu Lys
85 90 95

Lys Gln Asp Ser Val Phe Ala Thr Arg Pro Leu Thr Met Gly Thr Glu
100 105 110

Tyr Cys Ser Arg Gly Tyr Leu Thr Val Ala Val Glu Pro Gln Gly Glu
115 120 125

Gln Trp Lys Lys Met Arg Arg Val Val Ala Ser His Val Thr Ser Lys
130 135 140

Lys Ser Phe Gln Met Met Leu Gln Lys Arg Thr Glu Glu Ala Asp Asn
145 150 155 160

Leu Val Arg Tyr Ile Asn Asn Arg Ser Val Lys Asn Arg Gly Asn Ala
165 170 175

Phe Val Val Ile Asp Leu Arg Leu Ala Val Arg Gln Tyr Ser Gly Asn
180 185 190

Val Ala Arg Lys Met Met Phe Gly Ile Arg His Phe Gly Lys Gly Ser
195 200 205

Glu Asp Gly Ser Gly Pro Gly Leu Glu Glu Ile Glu His Val Glu Ser
210 215 220

Leu Phe Thr Val Leu Thr His Leu Tyr Ala Phe Ala Leu Ser Asp Tyr
225 230 235 240

Val Pro Trp Leu Arg Phe Leu Asp Leu Glu Gly His Glu Lys Val Val
245 250 255

Ser Asn Ala Met Arg Asn Val Ser Lys Tyr Asn Asp Pro Phe Val Asp
260 265 270

Glu Arg Leu Met Gln Trp Arg Asn Gly Lys Met Lys Glu Pro Gln Asp
275 280 285

Phe Leu Asp Met Phe Ile Ile Ala Lys Asp Thr Asp Gly Lys Pro Thr
290 295 300

Leu Ser Asp Glu Glu Ile Lys Ala Gln Val Thr Glu Leu Met Leu Ala
305 310 315 320

Thr Val Asp Asn Pro Ser Asn Ala Ala Glu Trp Gly Met Ala Glu Met
325 330 335

Ile Asn Glu Pro Ser Ile Met Gln Lys Ala Val Glu Glu Ile Asp Arg
340 345 350

Val Val Gly Lys Asp Arg Leu Val Ile Glu Ser Asp Leu Pro Asn Leu
355 360 365

Asn Tyr Val Lys Ala Cys Val Lys Glu Ala Phe Arg Leu His Pro Val
370 375 380

Ala Pro Phe Asn Leu Pro His Met Ser Thr Thr Asp Thr Val Val Asp
385 390 395 400

Gly Tyr Phe Ile Pro Lys Gly Ser His Val Leu Ile Ser Arg Met Gly
405 410 415

Ile Gly Arg Asn Pro Ser Val Trp Asp Lys Pro His Lys Phe Asp Pro
420 425 430

Glu Arg His Leu Ser Thr Asn Thr Cys Val Asp Leu Asn Glu Ser Asp
435 440 445

Leu Asn Ile Ile Ser Phe Ser Ala Gly Arg Arg Gly Cys Met Gly Val
450 455 460

Asp Ile Gly Ser Ala Met Thr Tyr Met Leu Leu Ala Arg Leu Ile Gln
465 470 475 480

Gly Phe Thr Trp Leu Pro Val Pro Gly Lys Asn Lys Ile Asp Ile Ser

485 490 495

Glu Ser Lys Asn Asp Leu Phe Met Ala Lys Pro Leu Tyr Ala Val Ala
500 505 510

Thr Pro Arg Leu Ala Pro His Val Tyr Pro Thr
515 520

<210> 40
<211> 1572
<212> DNA
<213> Arabidopsis thaliana

<400> 40
atgctcgcgt ttattatagg tttgcttctt cttgcattaa ctatgaagcg taaggagaag 60
aagaaaacca tgttaattag ccctacgaga aacctctctc tccctcccgg gccgaaatct 120
tggcctttaa tcggaaacct accggaaata ctagggagga acaaaccggt gttccggtgg 180
atacattctc tcatgaaaga actcaacacc gatattgcat gtatccgtct tgcgaatact 240
cacgtgatcc ccgtgacatc cccgagaatt gcaagagaga ttctgaagaa gcaagactcc 300
gttttcgcca ctagaccgct aacgatgggc acggagtact gcagccgcgg gtacttgacc 360
gttgcggtgg agccacaagg agagcagtgg aagaagatga ggagagtggt ggcatctcac 420
gtgacgagca agaagagctt ccaaatgatg ctacaaaaga gaaccgaaga ggctgataac 480
ttagtccggt acatcaataa ccgtagtgtc aaaaaccgtg gtaatgcttt tgtggttatt 540
gatttaaggc ttgcggtacg gcaatacagt ggaaatgtag ctcggaagat gatgtttggt 600
ataaggcatt ttggtaaagg aagtgaagat ggatcgggac cagggttgga agagattgaa 660
catgtggaat ctttgtttac ggttttaacc catctttacg cctttgcatt gtcagattat 720
gtcccgtggc taaggttctt ggacttggaa ggccatgaga aggttgtgag taacgcaatg 780
agaaatgtaa gtaagtataa cgaccctttt gttgatgaaa gactcatgca atggcgaaat 840
gggaagatga aagaacctca agattttctt gacatgttta taatagctaa agacactgac 900
gggaagccta ctctgtcgga cgaagagatc aaagcacaag tgacggaact aatgttggcg 960
acggttgata atccgtctaa cgcggcagag tggggtatgg cggagatgat taacgagccg 1020
agcatcatgc aaaaagccgt ggaagagatt gatagggtag ttggaaaaga ccgtcttgtc 1080
attgagtctg atctcccaaa tcttaactat gtgaaggctt gtgtgaaaga agcattccgg 1140
ttacaccccg tggcaccgtt caacctccct cacatgtcca ccactgatac tgtggtagac 1200
ggttatttca tccccaaggg aagccacgta ttgattagtc gtatggggat tgggagaaat 1260
cctagtgtgt gggacaagcc gcataagttc gaccctgaga gacatttgag cactaacaca 1320
tgtgtggatc taaacgagtc tgatctgaat ataatatcgt tcagtgcagg acgaagaggt 1380
tgtatgggtg tggacattgg gtcagccatg acgtacatgt tactggctcg gttgattcaa 1440
ggattcacgt ggttaccagt gcctggtaag aataagattg atatttcaga aagcaagaat 1500
gatcttttta tggcaaaacc attatacgcg gttgccacac ctcgtttagc tccacatgtg 1560
tatccaacct aa 1572

<210> 41
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer
A2F1

<400> 41
gtgcatatgc ttgactccac cccaatg 27

<210> 42
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer
A2R1

<400> 42
atgcattttt ctagtaatct ttacgctc 28

<210> 43
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer
A2F2

<400> 43
cgtgaattcc atatgctcgc gtttattata ggtttgc 37

<210> 44
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer
A2R2

<400> 44
cggaagctta ttaggttgga tacacatgt 29

<210> 45
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer
A2R3

<400> 45
cgtcacttgt gctttgatct cttc 24

<210> 46
<211> 24
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer
A2F3

<400> 46
gaactaatgt tggcgacggt tgat 24

<210> 47
<211> 57
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer
A2FX1

<400> 47
cgtgaattcc atatggctct gttattagca gtttttctcg cgtttattat aggtttg 57

<210> 48
<211> 57
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer
A2FX2

<400> 48
cgtgaattcc atatggctct gttattagca gtttttcttc ttcttgcatt aactatg 57

<210> 49
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer
A2R4

<400> 49
catctcgagt cttcttccac tgctctcctt 30

<210> 50
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

A2FX3

<400> 50
ttaatcggaa acctacc 17

<210> 51
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer 17AF

<400> 51
cgtgaattcc atatggctct gttattagct gtt 33

<210> 52
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer A1R

<400> 52
gggccacggc acgggacc 18

<210> 53
<211> 2702
<212> DNA
<213> Arabidopsis thaliana

<400> 53
ctcgagctca gtttcttctt cttcctcgta cttatcctcc tcagccaaac gatctctcac 60
cgtattctct agctgcactc cgtactgagc tccttttatc tcctttatca ccaccactct 120
tataaccttc tccatctccg ctgaaaaatg tataatagta agcagaggaa ccggttcaat 180
ttcgttggac acgtacttaa ccagattaat taagtaaacc ggagtttaac cagttgaatc 240
aaagtaaacc aaaataagaa gccaaaccaa ataatgtatt tattgaacca cgtagtctcc 300
atctaaacca gagaacccta attcaaattt tgatttgaaa acatggacta attaagatta 360
ccggaggcaa gggcgtcgaa gaagtcatca tcgccggcga gttcttttcc ggttttgcct 420
ttccagttat caagatgtgt tttaacatct gaaggatcta agtaaactcc gatcgcagtg 480
aacttcactt gaagaaagtg gatctcaatg tctgtgatcc ctataacaag aatatgaaca 540
atccatataa aattattgtt acctgcgatt ttgttatgta tcgcattata aggtatcaga 600
cattaagaaa gcaaaaaaga aataaaaacc ttggcccaga agagagagtg gcttggaagt 660
gatgatctgt ggaggaaaag gaacctcgtg aaccatgacc atctctgttc ccactgtttg 720
ataaacaaga acacacaaat cttaggaaaa aaacaaagca ttgaaaaaaa gacaatgaga 780
ataattgaaa cttgttagaa ctgaaaatct tactttagtg gataaacttg taataaaaaa 840
gaatgcaaag agtgtaagac ttactttcta atttatatta ttttgaatct gagagtgaag 900
aaatttataa atggcttggt gtactatttt acgatcttag agaaacaata tcgaaattgt 960
aaatgtgaat atctctctct atataataag ccagggactg gtggtaggta acataatttt 1020
gctaacgttc aaagcttgtg atctaaaaga cgacgtattc tttttatgaa ttcaattttt 1080
ttgctaccaa agcttgtgat ctcaattgtt tgttagcgac ccaccaggaa gccacgtgtt 1140

tggatcaagc actcagtcca caaccactca ttctacctaa caaatgaagg tatagaagta 1200
taataattaa aagagataga agaaagaatt gctatgatac agtaaaaaga gatcagatgt 1260
caaatgtgaa acaaagcgta cataaattag atacaaaatt agaagcagcc acatttctcc 1320
acaacggctc ttgaaatcag taacgtaaag taaactgatg atgacaaaga cccaaaaaaa 1380
aaaaaaaaaa aaaaaaagag aaataaagag tgtctttaaa gcagtaacgt ataaaaaccc 1440
tttttcgtct tctcttctat ctcgacctcc caaatcatga aaggatcaat tcatgactcc 1500
gctattacgg gtttagaggc tcagcttatg gcatcgaagc gaggactaat cgaagccgtg 1560
agtttgggga tcactcataa gtcatatctc aatctattgc gatcattatc acatttttga 1620
actggtaagt aagtgttact gctgcaaatc gaaactgact attgaaagct atgcccatct 1680
ttcgacacat aaactaagag ccaagtggga acaaaggatc gaagagacaa ctgaaagaga 1740
ttgggtaatg tgtgcaaagt gccaaaaatt ggcttcagca agtcatggta taatctctat 1800
tctctaatca caatctctag cttttcttaa ttagtcctta tgtaatttga ttatgtttta 1860
attcgcctcc taattaattt catggttgat ggatagtcgt gggtattcct tttgctacgc 1920
atgtcgagcc gaatggaagc tgctaggatt aaatttacag aagctgaatc aatttttaag 1980
tgggccaaat atttacagtt tttataagcc caaatctcca tgtccatatt gtttttaacg 2040
tggcgctacc taaaagggga taaagatttc ataaacagca ttaacaattt aacatcaaca 2100
agattttaaa gggataagga ttaaggaatc gtaagcaaat ttatccttag agattagatt 2160
tagacgaatt tggaaaagta aaaagttggt aattaaatag aaatgtactt aaaacacaac 2220
atgtaataca ttagacatat gagctgttga aaaatcgtgg tttttctaat gatggcgcta 2280
cctaaaaggg acaaggattt cataatgatg cattaccaat ttaacatcca caagatttat 2340
aagggataag gaataatcaa agaaaaaaac atgtcttaca tatgagctgt tgaaaaatcg 2400
tggattcatt taacattgtt ttcttcaaca tttaaagcac atttattttc catagattac 2460
acttaaacaa aagcatttgt ttcatggcta taaatagctt attcctcatc atagataaga 2520
aaaaaccttt tcgaactcaa ataatttctc caaattgaga tttaaaaaaa aaaatgcttg 2580
actccacccc aatgctcgcg tttattatag gtttgcttct tcttgcatta actatgaagc 2640
gtaaggagaa gaagaaaacc atgttaatta gccctacgag aaacctctct ctccctcccg 2700
gg 2702

<210> 54
<211> 541
<212> PRT
<213> Arabidopsis thaliana

<400> 54
Met Asn Thr Phe Thr Ser Asn Ser Ser Asp Leu Thr Thr Thr Ala Thr
1 5 10 15

Glu Thr Ser Ser Phe Ser Thr Leu Tyr Leu Leu Ser Thr Leu Gln Ala
20 25 30

Phe Val Ala Ile Thr Leu Val Met Leu Leu Lys Lys Leu Met Thr Asp
35 40 45

Pro Asn Lys Lys Lys Pro Tyr Leu Pro Pro Gly Pro Thr Gly Trp Pro
50 55 60

Ile Ile Gly Met Ile Pro Thr Met Leu Lys Ser Arg Pro Val Phe Arg
65 70 75 80

Trp Leu His Ser Ile Met Lys Gln Leu Asn Thr Glu Ile Ala Cys Val
85 90 95

Lys Leu Gly Asn Thr His Val Ile Thr Val Thr Cys Pro Lys Ile Ala
100 105 110

```
Arg Glu Ile Leu Lys Gln Gln Asp Ala Leu Phe Ala Ser Arg Pro Leu
        115                 120                 125

Thr Tyr Ala Gln Lys Ile Leu Ser Asn Gly Tyr Lys Thr Cys Val Ile
    130                 135                 140

Thr Pro Phe Gly Asp Gln Phe Lys Lys Met Arg Lys Val Val Met Thr
145                 150                 155                 160

Glu Leu Val Cys Pro Ala Arg His Arg Trp Leu His Gln Lys Arg Ser
                165                 170                 175

Glu Glu Asn Asp His Leu Thr Ala Trp Val Tyr Asn Met Val Lys Asn
            180                 185                 190

Ser Gly Ser Val Asp Phe Arg Phe Met Thr Arg His Tyr Cys Gly Asn
        195                 200                 205

Ala Ile Lys Lys Leu Met Phe Gly Thr Arg Thr Phe Ser Lys Asn Thr
    210                 215                 220

Ala Pro Asp Gly Gly Pro Thr Val Glu Asp Val Glu His Met Glu Ala
225                 230                 235                 240

Met Phe Glu Ala Leu Gly Phe Thr Phe Ala Phe Cys Ile Ser Asp Tyr
                245                 250                 255

Leu Pro Met Leu Thr Gly Leu Asp Leu Asn Gly His Glu Lys Ile Met
            260                 265                 270

Arg Glu Ser Ser Ala Ile Met Asp Lys Tyr His Asp Pro Ile Ile Asp
        275                 280                 285

Glu Arg Ile Lys Met Trp Arg Glu Gly Lys Arg Thr Gln Ile Glu Asp
    290                 295                 300

Phe Leu Asp Ile Phe Ile Ser Ile Lys Asp Glu Gln Gly Asn Pro Leu
305                 310                 315                 320

Leu Thr Ala Asp Glu Ile Lys Pro Thr Ile Lys Glu Leu Val Met Ala
                325                 330                 335

Ala Pro Asp Asn Pro Ser Asn Ala Val Glu Trp Ala Met Ala Glu Met
            340                 345                 350

Val Asn Lys Pro Glu Ile Leu Arg Lys Ala Met Glu Glu Ile Asp Arg
        355                 360                 365

Val Val Gly Lys Glu Arg Leu Val Gln Glu Ser Asp Ile Pro Lys Leu
    370                 375                 380

Asn Tyr Val Lys Ala Ile Leu Arg Glu Ala Phe Arg Leu His Pro Val
385                 390                 395                 400

Ala Ala Phe Asn Leu Pro His Val Ala Leu Ser Asp Thr Thr Val Ala
                405                 410                 415
```

```
Gly Tyr His Ile Pro Lys Gly Ser Gln Val Leu Leu Ser Arg Tyr Gly
            420                 425                 430

Leu Gly Arg Asn Pro Lys Val Trp Ala Asp Pro Leu Cys Phe Lys Pro
        435                 440                 445

Glu Arg His Leu Asn Glu Cys Ser Glu Val Thr Leu Thr Glu Asn Asp
    450                 455                 460

Leu Arg Phe Ile Ser Phe Ser Thr Gly Lys Arg Gly Cys Ala Ala Pro
465                 470                 475                 480

Ala Leu Gly Thr Ala Leu Thr Thr Met Met Leu Ala Arg Leu Leu Gln
                485                 490                 495

Gly Phe Thr Trp Lys Leu Pro Glu Asn Glu Thr Arg Val Glu Leu Met
            500                 505                 510

Glu Ser Ser His Asp Met Phe Leu Ala Lys Pro Leu Val Met Val Gly
        515                 520                 525

Asp Leu Arg Leu Pro Glu His Leu Tyr Pro Thr Val Lys
    530                 535                 540
```

<210> 55
<211> 1916
<212> DNA
<213> Arabidopsis thaliana

<400> 55

```
gtcgacccac gcgtccgcaa cagaaaccac aacaaaaact ttgagtcctc ttcttctcta 60
tacacaaaca tgaacacttt tacctcaaac tcttcggatc tcactaccac tgcaaccgaa 120
acatcgtcct ttagcacctt gtatctcctc tcaacacttc aagcttttgt ggctataacc 180
ttagtgatgc tactcaagaa attgatgacg gatcccaaca aaaagaaacc gtatctgcca 240
ccgggtccca caggatggcc gatcattgga atgattccga cgatgctaaa gagccggccc 300
gttttccggt ggctccacag catcatgaag cagctcaata ctgagatagc atgcgtgaag 360
ttaggaaaca ctcatgtgat caccgtcacg tgccctaaga tagcacgtga gatactcaag 420
caacaagacg ctctcttcgc gtcgaggcct ttaacttacg ctcagaagat cctctctaac 480
ggctacaaaa cctgcgtgat cactcccttt ggtgaccaat tcaagaaaat gaggaaagtt 540
gtgatgacgg aactcgtatg tccagcgaga cacaggtggc tccaccagaa gagatcagaa 600
gaaaacgatc atttaaccgc ttgggtatac aacatggtta agaactcggg ctctgtcgat 660
ttccggttca tgactaggca ttactgtgga aatgcaatca agaagcttat gttcgggacg 720
agaacgttct ctaagaacac tgcacctgac ggtggaccca ccgtagaaga tgtagagcac 780
atggaagcaa tgtttgaagc attagggttt accttcgctt tttgcatctc tgattatctg 840
ccgatgctca ctggacttga tcttaacggt cacgagaaga ttatgagaga atcaagtgcg 900
attatggaca agtatcatga cccaatcatc gacgagagga tcaagatgtg gagagaagga 960
aagagaactc aaatcgaaga ttttcttgat attttcatct ctatcaaaga cgaacaaggc 1020
aacccattgc ttaccgccga tgaaatcaaa cccaccatta aggagcttgt aatggcggcg 1080
ccagacaatc catcaaacgc cgtggaatgg gccatggcgg agatggtgaa caaaccggag 1140
attctccgta aagcaatgga agagatcgac agagtcgtcg ggaaagagag actcgttcaa 1200
gaatccgaca tcccaaaact aaactacgtc aaagctatcc tccgcgaagc tttccgtctc 1260
catcccgtcg ccgccttcaa cctcccccac gtggcacttt ctgacacaac cgtcgccgga 1320
tatcacatcc ctaaaggaag tcaagtcctt cttagccgat atgggctggg ccgtaaccca 1380
aaagtttggg ccgacccact ttgctttaaa ccggagagac atctcaacga atgctccgaa 1440
```

gttactttga ccgagaacga tctccggttt atctcgttca gtaccgggaa aagaggttgt 1500
gcggctccgg cgctaggaac ggcgttgacc acgatgatgc tcgcgagact tcttcaaggt 1560
ttcacttgga agctacctga gaatgagaca cgtgtcgagc tgatggagtc tagtcacgat 1620
atgtttctgg ctaaaccgtt ggttatggtc ggtgacctta gattgccgga gcatctctac 1680
ccgacggtga agtgagatga gacgacgccg tatatatttt atgaaactac ttttatataa 1740
tcgcccaacc aagtttggtc aattccggtt accagaagat aattggtcaa attgtgaaca 1800
aacttgtgtg ttggtttctt ggttcttttt gggacacttg aattgtgtct cctttacctc 1860
ttcttttgtt gttttcaata aaaactttta ttaccatttc aaaaaaaaaa aaaaaa 1916

<210> 56
<211> 1974
<212> DNA
<213> Arabidopsis thaliana

<400> 56
atgaacactt ttacctcaaa ctcttcggat ctcactacca ctgcaaccga aacatcgtcc 60
tttagcacct tgtatctcct ctcaacactt caagcttttg tggctataac cttagtgatg 120
ctactcaaga aattgatgac ggatcccaac aaaaagaaac cgtatctgcc accgggtccc 180
acaggatggc cgatcattgg aatgattccg acgatgctaa agagccggcc cgttttccgg 240
tggctccaca gcatcatgaa gcagctcaat actgagatag catgcgtgaa gttaggaaac 300
actcatgtga tcaccgtcac gtgccctaag atagcacgtg agatactcaa gcaacaagac 360
gctctcttcg cgtcgaggcc tttaacttac gctcagaaga tcctctctaa cggctacaaa 420
acctgcgtga tcactccctt tggtgaccaa ttcaagaaaa tgaggaaagt tgtgatgacg 480
gaactcgtat gtccagcgag acacaggtgg ctccaccaga agagatcaga agaaaacgat 540
catttaaccg cttgggtata caacatggtt aagaactcgg gctctgtcga tttccggttc 600
atgactaggc attactgtgg aaatgcaatc aagaagctta tgttcgggac gagaacgttc 660
tctaagaaca ctgcacctga cggtggaccc accgtagaag atgtagagca catggaagca 720
atgtttgaag cattagggtt taccttcgct ttttgcatct ctgattatct gccgatgctc 780
actggacttg atcttaacgg tcacgagaag attatgagag aatcaagtgc gattatggac 840
aagtatcatg acccaatcat cgacgagagg atcaagatgt ggagagaagg aaagagaact 900
caaatcgaag attttcttga tattttcatc tctatcaaag acgaacaagg caacccattg 960
cttaccgccg atgaaatcaa acccaccatt aaggtattta tcacgttcct ttcatataag 1020
gtttcgatcg taaaaatatc aaaagaacaa tttttgttaa attttatttg agaaagcatg 1080
catatcaaat ttatttacac atactaacat tttgattcat aaaacattta taaaagaaga 1140
aagaaacatt ttgtggtaaa agttgattag ttacaatatt tgtttttttt ttgctaaaca 1200
tgggctactt ttttgtttgt ctcttttgat tactttggtc aaagacagat gcatgcaact 1260
taattgtatt tatttttatg ttatacaaaa attaaagatc caaaattaat aaaagctggt 1320
atatatgttt ataatgaata ggagcttgta atggcggcgc cagacaatcc atcaaacgcc 1380
gtggaatggg ccatggcgga gatggtgaac aaaccggaga ttctccgtaa agcaatggaa 1440
gagatcgaca gagtcgtcgg gaaagagaga ctcgttcaag aatccgacat cccaaaacta 1500
aactacgtca aagctatcct ccgcgaagct ttccgtctcc atcccgtcgc cgccttcaac 1560
ctcccccacg tggcactttc tgacacaacc gtcgccggat atcacatccc taaaggaagt 1620
caagtccttc ttagccgata tgggctgggc cgtaacccaa aagtttgggc cgacccactt 1680
tgctttaaac cggagagaca tctcaacgaa tgctccgaag ttactttgac cgagaacgat 1740
ctccggttta tctcgttcag taccgggaaa agaggttgtg cggctccggc gctaggaacg 1800
gcgttgacca cgatgatgct cgcgagactt cttcaaggtt tcacttggaa gctacctgag 1860
aatgagacac gtgtcgagct gatggagtct agtcacgata tgtttctggc taaaccgttg 1920
gttatggtcg gtgaccttag attgccggag catctctacc cgacggtgaa gtga 1974

<210> 57
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer T7

<400> 57
aatacgactc actatag 17

<210> 58
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer EST3

<400> 58
gctaggatcc atgttgtata cccaag 26

<210> 59
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer EST6

<400> 59
cgggcccgtt ttccggtggc 20

<210> 60
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer EST7A

<400> 60
ggtcaccaaa gggagtgatc acgc 24

<210> 61
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer 5´
´native´ sense

<400> 61
atcgtcagtc gaccatatga acacttttac ctcaaactct tcgg 44

<210> 62
<211> 68
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer 5´
´bovine´ sense

<400> 62
atcgtcagtc gaccatatgg ctctgttatt agcagttttt acatcgtcct ttagcacctt 60
gtatctcc 68

<210> 63
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer 3´
´end´ antisense

<400> 63
actgctagaa ttcgacgtca ttacttcacc gtcgggtaga gatgc 45

<210> 64
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer
CYP79B2.2

<400> 64
ggaattcatg aacactttta cctca 25

<210> 65
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer B2SB

<400> 65
ttgtctagat cacttcaccg tcgggta 27

<210> 66
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer B2AF

<400> 66
ggcctcgaga tgaacacttt tacctca 27

<210> 67
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer B2AB

<400> 67
ttggaattcc ttcaccgtcg ggtagag 27

<210> 68
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer Xba I

<400> 68
gtaccatcta gattcatgtt tgtgtataga g 31

<210> 69
<211> 2361
<212> DNA
<213> Arabidopsis thaliana

<400> 69
gaattcattg atctggtctt gctaaaaact ttaaaattga tgagttcaac atcttcaaat 60
gcatgataac gggtccaacg gaaattgact ttttttcat gctcctgata tataataata 120
tctaacgatt acgggttcca ctaattgtca ttactcatta acattcctat ttaaaagttg 180
tgatagtttt agggttttac gtagtcgtgt catatagcga ttaactacgt acttgtagat 240
ttatcaatta cttctgttgt ttacgagaac ctaaaaaaaa gaagcagatg cctagtttat 300
agagcacgtg tactgtcttg aaaacttagg taggttggta aggttaccaa aagaccttaa 360
aggaatataa agttactaat taacttaagt aaagttggta ttgcttatat attgcaaagt 420
attacaaacc aatcccctct gtatattgtt ttaaaccata gatttttta caattaagtt 480
tatgatcaat caattatttc accatttcta ttaaattatg taaaaagaaa aggatatata 540
tatatatata taattaaata agaataaatc aaaataccga aatttttat tatccattct 600
ttgtggacat cgcccctaat atataaaaaa aaaaaacttt cgtataactg atttatattt 660
ttttgtaaaa acttaaagga agcctaagaa atatcttgtg atatttttga caaaatgtgg 720
tatatatctt tttataatat catttataaa gaaaatattg attacatggt gaaaaacatt 780
ttgctagcga tcaacaaaat taaataggca catgttaact gatctcatac gaccttgaaa 840
ttttaatctt tgtgtcgaga gaccgatctt tatgcaaatt atgaaactac acatggttta 900
tgcacggaag atcacattgc atgtatacca tattataaac caaaaatgat caagaagaag 960
gcgaaaacat ttgggtaaat tttaaatttc gatcatgcga ttttttagct catcatcaac 1020
agacaagaaa ctatcttttg tactgtaaat actaaataca aaataaaatc ttcatcattt 1080

tttgcatgtg tcaaataaat tacgcgaact ttttttttt atcgactatt aatagagaaa 1140
cctgttttat ttgccttgat ttggaaaaat ggagaaattg acttaagact tagtctcggt 1200
cacatcggca acaacggagc ttaaacggcg tccgcaacat ggaaactcaa gccacgaatc 1260
tgatatattg actatagaag tagtaagtaa ctttgactcg tcccacatca gtttcaattt 1320
ccacgagggt atttggcagg tgaactctct acgtacccaa aacataatgg ctattttatt 1380
tcataactga tatttagcaa ttaattattc gtccttttta aaccaatttc tatagttggg 1440
aaaataatca atttttacac tttcaatgta tacgttacag attttttttt attagtcatg 1500
cacatatttt caatttttac actttcaatg taaacaatcg attcttaatt gttaaaaata 1560
ggtttacgta aggaattaaa gatttgttta aaatatgttc cggccggtct aataatttac 1620
ttgacgttaa tttcttaaac acttttagat aggaggcttt gtttatccca aatgattttg 1680
taccactgcg acaatactag ctagacataa aatgttaata aatttttatt aagtaatata 1740
atcgaagtat tagatcaatg tagtagacag ttaggttaac taaaacaaga gtaaacactt 1800
ttttttttct tttcaggata ggtaaaacaa atttcacact attttgcgta tttccttaaa 1860
tttgttgttc gttttctcag caaagatgaa tattttgttt catagtaatt cacaagtata 1920
aactcgccag aactcctcaa acagtgaaat ataatatagc ttttaactgt ttttcggctg 1980
gaccgggttt ttaagtgcat atataacacg aggaattttg gcaggtcacc aacaaaactt 2040
ttaaaaatat taaaaattcc catcaagaat agaaattaat aaacaatgat atctctaata 2100
atatagatat tttgaaacgt taggaataat cgtaataatg ttcaacgttg gtggtggtac 2160
tcaagatgga ccctccctcc cacattttcc tcactccttc gtaagtcctt tccacgcata 2220
agggtattat agtcatttca cataaactaa cgactactag acttgtatat aaataggaag 2280
gtgaagctct ctctttatcc atgcagagac aacagaaacc acaacaaaaa ctttgagtcc 2340
tcttcttctc tatacacaaa c 2361

<210> 70
<211> 540
<212> PRT
<213> Brassica napus

<400> 70
Met Asn Thr Phe Thr Ser Asn Ser Ser Asp Leu Thr Ser Thr Thr Thr
1 5 10 15

Gln Thr Ser Pro Phe Ser Asn Met Tyr Leu Leu Thr Thr Leu Gln Ala
20 25 30

Phe Ala Ala Ile Thr Leu Val Met Leu Leu Lys Lys Val Phe Thr Thr
35 40 45

Asp Lys Lys Lys Leu Ser Leu Pro Pro Gly Pro Thr Gly Trp Pro Ile
50 55 60

Ile Gly Met Val Pro Thr Met Leu Lys Ser Arg Pro Val Phe Arg Trp
65 70 75 80

Leu His Ser Ile Met Lys Gln Leu Asn Thr Glu Ile Ala Cys Val Arg
85 90 95

Leu Gly Asn Thr His Val Ile Thr Val Thr Cys Pro Lys Ile Ala Arg
100 105 110

Glu Ile Leu Lys Gln Gln Asp Ala Leu Phe Ala Ser Arg Pro Met Thr
115 120 125

Tyr Ala Gln Asn Val Leu Ser Asn Gly Tyr Lys Thr Cys Val Ile Thr
130 135 140

```
Pro Phe Gly Glu Gln Phe Lys Lys Met Arg Lys Val Val Met Thr Glu
145                 150                 155                 160

Leu Val Cys Pro Ala Arg His Arg Trp Leu His Gln Lys Arg Ala Glu
                165                 170                 175

Glu Asn Asp His Leu Thr Ala Trp Val Tyr Asn Leu Val Lys Asn Ser
            180                 185                 190

Gly Ser Val Asp Phe Arg Phe Val Thr Arg His Tyr Cys Gly Asn Ala
        195                 200                 205

Ile Lys Lys Leu Met Phe Gly Thr Arg Thr Phe Ser Glu Asn Thr Ala
    210                 215                 220

Pro Asp Gly Gly Pro Thr Ala Glu Asp Ile Glu His Met Glu Ala Met
225                 230                 235                 240

Phe Glu Ala Leu Gly Phe Thr Phe Ser Phe Cys Ile Ser Asp Tyr Leu
                245                 250                 255

Pro Met Leu Thr Gly Leu Asp Leu Asn Gly His Glu Lys Ile Met Arg
            260                 265                 270

Asp Ser Ser Ala Ile Met Asp Lys Tyr His Asp Pro Ile Val Asp Ala
        275                 280                 285

Arg Ile Lys Met Trp Arg Glu Gly Lys Arg Thr Gln Ile Glu Asp Phe
    290                 295                 300

Leu Asp Ile Phe Ile Ser Ile Lys Asp Glu Gln Gly Asn Pro Leu Leu
305                 310                 315                 320

Thr Ala Asp Glu Ile Lys Pro Thr Ile Lys Glu Leu Val Met Ala Ala
                325                 330                 335

Pro Asp Asn Pro Ser Asn Ala Val Glu Trp Ala Met Ala Glu Met Val
            340                 345                 350

Asn Lys Pro Glu Ile Leu His Lys Ala Met Glu Glu Ile Asp Arg Val
        355                 360                 365

Val Gly Lys Glu Arg Leu Val Gln Glu Ser Asp Ile Pro Lys Leu Asn
    370                 375                 380

Tyr Val Lys Ala Ile Leu Arg Glu Ala Phe Arg Leu His Pro Val Ala
385                 390                 395                 400

Ala Phe Asn Leu Pro His Val Ala Leu Ser Asp Ala Thr Val Ala Gly
                405                 410                 415

Tyr His Ile Pro Lys Gly Ser Gln Val Leu Leu Ser Arg Tyr Gly Leu
            420                 425                 430

Gly Arg Asn Pro Lys Val Trp Ala Asp Pro Leu Ser Phe Lys Pro Glu
```

435 440 445

Arg His Leu Asn Glu Cys Ser Glu Val Thr Leu Thr Glu Asn Asp Leu
450 455 460

Arg Phe Ile Ser Phe Ser Thr Gly Lys Arg Gly Cys Ala Ala Pro Ala
465 470 475 480

Leu Gly Thr Ala Leu Thr Thr Met Met Leu Ala Arg Leu Leu Gln Gly
485 490 495

Phe Thr Trp Lys Leu Pro Glu Asn Glu Thr Arg Val Glu Leu Met Glu
500 505 510

Ser Ser His Asp Met Phe Leu Ala Lys Pro Leu Val Met Val Gly Glu
515 520 525

Leu Arg Leu Pro Glu His Leu Tyr Pro Thr Val Lys
530 535 540

<210> 71
<211> 1913
<212> DNA
<213> Brassica napus

<400> 71
tggagctcca ccgcggtggc ggccgctcta gaactagtgg atcccccggg ctgcaggaat 60
tcgcggccgc gtcgactttg attcttcttc tctgctctct ctctctctac tcgaaaacat 120
gaacaccttt acctcaaact cttcggatct cacttccact acaacgcaaa cgtctccgtt 180
cagcaacatg tatctcctca caacgctcca ggcctttgcg gctataacct tggtgatgct 240
tctcaagaaa gtcttcacga cggataaaaa gaaattgtct ctcccgccgg gtcccaccgg 300
atggccgatc atcggaatgg ttccaacgat gctaaagagc cgtcccgttt tccggtggct 360
ccacagcatc atgaagcagc taaacaccga gatagcctgc gtgaggctag gaaacactca 420
cgtgatcacc gtcacatgcc cgaagatagc acgtgagata ctcaagcaac aagacgctct 480
cttcgcctcg agacccatga cttacgcaca gaatgtcctc tctaacggat acaaaacatg 540
cgtgatcact cccttcggtg aacaattcaa gaaaatgagg aaagtcgtga tgactgaact 600
cgtttgtccc gcgaggcaca ggtggcttca ccagaagaga gctgaagaga acgaccattt 660
aaccgcttgg gtatacaact tggtcaagaa ctctggctca gtcgattttc ggtttgtcac 720
gaggcattac tgtggaaatg ctatcaagaa gcttatgttc gggacaagaa cgttctctga 780
aaacaccgca cctgacggtg gaccaaccgc tgaggatatc gagcatatgg aagctatgtt 840
cgaagcatta gggtttactt tctccttttg tatctctgat tatctaccta tgctcactgg 900
acttgatctt aacggccacg agaagatcat gagggattcg agtgctatta tggacaagta 960
tcacgatcct atcgtcgatg caaggatcaa gatgtggaga gaaggaaaga gaactcaaat 1020
cgaggatttt ctagacattt ttatttctat caaggatgaa caaggcaacc cattgcttac 1080
cgccgatgaa atcaaaccca ccattaagga acttgtaatg gcggcgccag acaatccatc 1140
aaacgctgtc gagtgggcca tggcggagat ggtgaacaaa ccggagatac tccataaagc 1200
aatggaagaa atagacagag ttgtcggaaa agaaagactt gtccaagaat ccgacattcc 1260
aaaattaaat tacgtcaaag ctatcctccg tgaagccttc cgcctccatc ccgtagcggc 1320
ctttaacctc ccacacgtgg cactttccga cgcaaccgtc gccgggtatc acatccctaa 1380
aggaagtcaa gtccttctca gtcgatatgg gctgggccgt aacccgaaag tttgggctga 1440
ccccttgagc tttaaaccgg agagacatct caacgaatgc tcggaagtta ctttgacgga 1500
gaacgatctc cggtttatct cgtttagtac cgggaaaaga ggttgtgctg ctccggcttt 1560
aggtacggcg ttgaccacga tgatgctcgc gagacttctt caaggtttca cttggaagct 1620
gccggagaat gagacacgcg ttgagctgat ggagtctagc catgatatgt ttttggctaa 1680
accattggtt atggtcggtg agttgagact cccagagcat ctttacccga cggtgaagta 1740

agaataaaac gacggcgtat atattttatt aaataacttc tacgtactta tgtaattaac 1800
cacagagttt ggtcggtttc tccggttacc agaagataat cggttaatat atgaacaaac 1860
ttgtgcttgg ttttggtaaa aaaaaaaaaa aaaaaaaact cgaggggggg ccc 1913

<210> 72
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer EST1

<400> 72
tccatgtgct ctacatct 18

<210> 73
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer EST2

<400> 73
gacggaactc gtatgtcc 18

<210> 74
<211> 537
<212> PRT
<213> Arabidopsis thaliana

<400> 74
Met Ser Phe Thr Thr Ser Leu Pro Tyr Pro Phe His Ile Leu Leu Val
1 5 10 15

Phe Ile Leu Ser Met Ala Ser Ile Thr Leu Leu Gly Arg Ile Leu Ser
20 25 30

Arg Pro Thr Lys Thr Lys Asp Arg Ser Cys Gln Leu Pro Pro Gly Pro
35 40 45

Pro Gly Trp Pro Ile Leu Gly Asn Leu Pro Glu Leu Phe Met Thr Arg
50 55 60

Pro Arg Ser Lys Tyr Phe Arg Leu Ala Met Lys Glu Leu Lys Thr Asp
65 70 75 80

Ile Ala Cys Phe Asn Phe Ala Gly Ile Arg Ala Ile Thr Ile Asn Ser
85 90 95

Asp Glu Ile Ala Arg Glu Ala Phe Arg Glu Arg Asp Ala Asp Leu Ala
100 105 110

Asp Arg Pro Gln Leu Phe Ile Met Glu Thr Ile Gly Asp Asn Tyr Lys
115 120 125

Ser Met Gly Ile Ser Pro Tyr Gly Glu Gln Phe Met Lys Met Lys Arg
130 135 140

Val Ile Thr Thr Glu Ile Met Ser Val Lys Thr Leu Lys Met Leu Glu
145 150 155 160

Ala Ala Arg Thr Ile Glu Ala Asp Asn Leu Ile Ala Tyr Val His Ser
165 170 175

Met Tyr Gln Arg Ser Glu Thr Val Asp Val Arg Glu Leu Ser Arg Val
180 185 190

Tyr Gly Tyr Ala Val Thr Met Arg Met Leu Phe Gly Arg Arg His Val
195 200 205

Thr Lys Glu Asn Val Phe Ser Asp Asp Gly Arg Leu Gly Asn Ala Glu
210 215 220

Lys His His Leu Glu Val Ile Phe Asn Thr Leu Asn Cys Leu Pro Ser
225 230 235 240

Phe Ser Pro Ala Asp Tyr Val Glu Arg Trp Leu Arg Gly Trp Asn Val
245 250 255

Asp Gly Gln Glu Lys Arg Val Thr Glu Asn Cys Asn Ile Val Arg Ser
260 265 270

Tyr Asn Asn Pro Ile Ile Asp Glu Arg Val Gln Leu Trp Arg Glu Glu
275 280 285

Gly Gly Lys Ala Ala Val Glu Asp Trp Leu Asp Thr Phe Ile Thr Leu
290 295 300

Lys Asp Gln Asn Gly Lys Tyr Leu Val Thr Pro Asp Glu Ile Lys Ala
305 310 315 320

Gln Cys Val Glu Phe Cys Ile Ala Ala Ile Asp Asn Pro Ala Asn Asn
325 330 335

Met Glu Trp Thr Leu Gly Glu Met Leu Lys Asn Pro Glu Ile Leu Arg
340 345 350

Lys Ala Leu Lys Glu Leu Asp Glu Val Val Gly Arg Asp Arg Leu Val
355 360 365

Gln Glu Ser Asp Ile Pro Asn Leu Asn Tyr Leu Lys Ala Cys Cys Arg
370 375 380

Glu Thr Phe Arg Ile His Pro Ser Ala His Tyr Val Pro Ser His Leu
385 390 395 400

Ala Arg Gln Asp Thr Thr Leu Gly Gly Tyr Phe Ile Pro Lys Gly Ser
405 410 415

```
His Ile His Val Cys Arg Pro Gly Leu Gly Arg Asn Pro Lys Ile Trp
            420                 425                 430

Lys Asp Pro Leu Val Tyr Lys Pro Glu Arg His Leu Gln Gly Asp Gly
        435                 440                 445

Ile Thr Lys Glu Val Thr Leu Val Glu Thr Glu Met Arg Phe Val Ser
    450                 455                 460

Phe Ser Thr Gly Arg Arg Gly Cys Ile Gly Val Lys Val Gly Thr Ile
465                 470                 475                 480

Met Met Val Met Leu Leu Ala Arg Phe Leu Gln Gly Phe Asn Trp Lys
                485                 490                 495

Leu His Gln Asp Phe Gly Pro Leu Ser Leu Glu Glu Asp Asp Ala Ser
            500                 505                 510

Leu Leu Met Ala Lys Pro Leu His Leu Ser Val Glu Pro Arg Leu Ala
        515                 520                 525

Pro Asn Leu Tyr Pro Lys Phe Arg Pro
    530                 535
```

<210> 75
<211> 1614
<212> DNA
<213> Arabidopsis thaliana

<400> 75

```
atgagcttta ccacatcatt accataccct tttcacatcc tactagtctt tatcctctcc 60
atggcatcaa tcactctact gggtcgaata ctctcaaggc ccaccaaaac caaagaccga 120
tcttgccagc ttcctcctgg cccaccagga tggcccatcc tcggcaatct acccgaacta 180
ttcatgactc gtcctaggtc caaatatttc cgccttgcca tgaaagagct aaaaacagat 240
atagcatgtt tcaactttgc cggcatccgt gccatcacca taaactccga cgagatcgct 300
agagaagcgt ttagagagcg agacgcagat ttggcagacc ggcctcaact tttcatcatg 360
gagacaatcg gagacaatta caaatcaatg ggaatttcac cgtacggtga acaattcatg 420
aagatgaaaa gagtgatcac aacggaaatt atgtccgtta agacgttgaa aatgttggag 480
gctgcaagaa ccatcgaagc ggataatctc atagcttacg ttcactccat gtatcaacgg 540
tccgagacgg tcgatgttag agagctctcg agggtttatg gttacgcagt gaccatgcga 600
atgttgtttg gaaggagaca tgttacgaaa gaaaacgtgt tttctgatga tggaagacta 660
ggaaacgccg aaaaacatca tcttgaggtg attttcaaca ctcttaactg tttaccgagt 720
tttagtccag cggattacgt ggaacgatgg ttgagaggtt ggaatgttga tggtcaagag 780
aagagggtga cagagaactg taacattgtt cgtagttaca acaatcccat aatcgacgag 840
agggtccagt tgtggaggga agaaggtggt aaggctgctg ttgaagattg gcttgatacg 900
ttcattaccc taaaagatca aaacggaaag tacttggtca caccagacga aatcaaagct 960
caatgcgtag aattttgtat agcagcgatt gataatccgg caaataacat ggagtggaca 1020
cttggggaaa tgttaaagaa cccggagatt cttagaaaag ctctgaagga gttggatgaa 1080
gtagttggaa gagacaggct tgtgcaagaa tcagacatac caaatctaaa ctacttaaaa 1140
gcttgttgta gagaaacatt cagaattcac ccaagtgctc attatgtccc ttcccatctt 1200
gcgcgtcaag ataccaccct tgggggttat ttcattccca aaggtagcca cattcatgta 1260
tgccgccctg gactaggtcg taaccctaaa atatggaaag atccattggt atacaaaccg 1320
gagcgtcacc tccaaggaga cggaatcaca aaagaggtta ctctggtgga aacagagatg 1380
cgttttgtct cgtttagcac cggtcgacgt ggctgcatcg gtgttaaagt cgggacgatc 1440
```

atgatggtta tgttgttggc taggtttctt caagggttta actggaaact ccatcaagat 1500
tttggaccgt taagcctcga ggaagatgat gcatcattgc ttatggctaa acctcttcac 1560
ttgtccgttg agccacgctt ggcaccaaac ctttatccaa agttccgtcc ttaa 1614

<210> 76
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer sequence

<400> 76
ctctagattc gaacatatgg ctagctttac aacatcatta cc 42

<210> 77
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer sequence

<400> 77
cgggatcctt aaggacggaa ctttggata 29

<210> 78
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer sequence

<400> 78
aactgcagca tgatgagctt taccacatc 29

<210> 79
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer sequence

<400> 79
cgggatcctt aatggtggtg atgaggacgg aactttggat aa 42

<210> 80
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer sequence

<400> 80
aaagctcaat gcgtagaat 19

<210> 81
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer sequence

<400> 81
tttttagaca ccatcttgtt ttcttcttc 29

<210> 82
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer sequence

<400> 82
tgtagcggcg cattaagc 18

<210> 83
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer sequence

<400> 83
caaaagaata gaccgagata ggg 23

<210> 84
<211> 535
<212> PRT

<213> Arabidopsis thaliana

<400> 84

```
Met Lys Ile Ser Phe Asn Thr Cys Phe Gln Ile Leu Leu Gly Phe Ile
 1               5                  10                  15

Val Phe Ile Ala Ser Ile Thr Leu Leu Gly Arg Ile Phe Ser Arg Pro
            20                  25                  30

Ser Lys Thr Lys Asp Arg Cys Arg Gln Leu Pro Pro Gly Arg Pro Gly
        35                  40                  45

Trp Pro Ile Leu Gly Asn Leu Pro Glu Leu Ile Met Thr Arg Pro Arg
    50                  55                  60

Ser Lys Tyr Phe His Leu Ala Met Lys Glu Leu Lys Thr Asp Ile Ala
65                  70                  75                  80

Cys Phe Asn Phe Ala Gly Thr His Thr Ile Thr Ile Asn Ser Asp Glu
                85                  90                  95

Ile Ala Arg Glu Ala Phe Arg Glu Arg Asp Ala Asp Leu Ala Asp Arg
            100                 105                 110

Pro Gln Leu Ser Ile Val Glu Ser Ile Gly Asp Asn Tyr Lys Thr Met
        115                 120                 125

Gly Thr Ser Ser Tyr Gly Glu His Phe Met Lys Met Lys Lys Val Ile
    130                 135                 140

Thr Thr Glu Ile Met Ser Val Lys Thr Leu Asn Met Leu Glu Ala Ala
145                 150                 155                 160

Arg Thr Ile Glu Ala Asp Asn Leu Ile Ala Tyr Ile His Ser Met Tyr
                165                 170                 175

Gln Arg Ser Glu Thr Val Asp Val Arg Glu Leu Ser Arg Val Tyr Gly
            180                 185                 190

Tyr Ala Val Thr Met Arg Met Leu Phe Gly Arg Arg His Val Thr Lys
        195                 200                 205

Glu Asn Met Phe Ser Asp Asp Gly Arg Leu Gly Lys Ala Glu Lys His
    210                 215                 220

His Leu Glu Val Ile Phe Asn Thr Leu Asn Cys Leu Pro Gly Phe Ser
225                 230                 235                 240

Pro Val Asp Tyr Val Asp Arg Trp Leu Gly Gly Trp Asn Ile Asp Gly
                245                 250                 255

Glu Glu Glu Arg Ala Lys Val Asn Val Asn Leu Val Arg Ser Tyr Asn
            260                 265                 270

Asn Pro Ile Ile Asp Glu Arg Val Glu Ile Trp Arg Glu Lys Gly Gly
        275                 280                 285
```

```
Lys Ala Ala Val Glu Asp Trp Leu Asp Thr Phe Ile Thr Leu Lys Asp
    290                 295                 300

Gln Asn Gly Asn Tyr Leu Val Thr Pro Asp Glu Ile Lys Ala Gln Cys
305                 310                 315                 320

Val Glu Phe Cys Ile Ala Ala Ile Asp Asn Pro Ala Asn Asn Met Glu
                325                 330                 335

Trp Thr Leu Gly Glu Met Leu Lys Asn Pro Glu Ile Leu Arg Lys Ala
            340                 345                 350

Leu Lys Glu Leu Asp Glu Val Val Gly Lys Asp Arg Leu Val Gln Glu
        355                 360                 365

Ser Asp Ile Arg Asn Leu Asn Tyr Leu Lys Ala Cys Cys Arg Glu Thr
    370                 375                 380

Phe Arg Ile His Pro Ser Ala His Tyr Val Pro Pro His Val Ala Arg
385                 390                 395                 400

Gln Asp Thr Thr Leu Gly Gly Tyr Phe Ile Pro Lys Gly Ser His Ile
                405                 410                 415

His Val Cys Arg Pro Gly Leu Gly Arg Asn Pro Lys Ile Trp Lys Asp
            420                 425                 430

Pro Leu Ala Tyr Glu Pro Glu Arg His Leu Gln Gly Asp Gly Ile Thr
        435                 440                 445

Lys Glu Val Thr Leu Val Glu Thr Glu Met Arg Phe Val Ser Phe Ser
    450                 455                 460

Thr Gly Arg Arg Gly Cys Val Gly Val Lys Val Gly Thr Ile Met Met
465                 470                 475                 480

Ala Met Met Leu Ala Arg Phe Leu Gln Gly Phe Asn Trp Lys Leu His
                485                 490                 495

Arg Asp Phe Gly Pro Leu Ser Leu Glu Glu Asp Asp Ala Ser Leu Leu
            500                 505                 510

Met Ala Lys Pro Leu Leu Leu Ser Val Glu Pro Arg Leu Ala Ser Asn
        515                 520                 525

Leu Tyr Pro Lys Phe Arg Pro
    530                 535
```

<210> 85
<211> 1608
<212> DNA
<213> Arabidopsis thaliana

<400> 85

atgaagatta gctttaacac atgctttcaa atcttactag gatttatcgt cttcatcgca 60
tcaatcactt tactaggtcg aatattctca aggccttcca aaaccaaaga ccggtgtcgc 120
cagcttcctc ctggccgacc aggatggccc atcctcggca atctacccga actaatcatg 180
actcgtccta ggtccaaata tttccacctt gccatgaaag agctaaaaac ggatatcgca 240
tgtttcaact ttgccggaac ccacaccatc accataaact ccgacgagat cgctagagaa 300
gcttttagag agcgagacgc agatttggca gaccggcctc aactttccat cgtagagtcc 360
attggagaca attacaaaac aatgggaacc tcatcgtacg gtgaacattt catgaagatg 420
aaaaaagtga tcacaacgga aattatgtcc gttaaaacgt tgaatatgtt ggaagctgcg 480
agaaccatcg aagcggataa tctcattgct tacattcact cgatgtatca acggtcggag 540
acggtcgacg ttagagaact ttcgagagtt tatggttacg cagtgaccat gagaatgttg 600
tttggaagga gacatgtcac gaaagaaaac atgttttcgg atgatgggag actaggaaaa 660
gccgaaaaac atcatcttga ggtgattttc aacactctaa actgtttgcc aggttttagt 720
cccgtggatt acgtggaccg atggttaggt ggttggaata ttgatggtga agaggagaga 780
gcgaaagtga atgttaatct tgttcgtagt tacaacaatc ccataataga cgagagggtc 840
gaaatttgga gggaaaaagg tggtaaggct gctgtggaag attggcttga tacgttcatt 900
acgctaaaag atcaaaacgg aaactacttg gttacgccag acgaaatcaa agctcaatgc 960
gtcgaatttt gtatagcagc gatcgataat ccggcaaata acatggagtg gacacttggg 1020
gaaatgttaa agaacccgga gattcttaga aaagctctga aggagttgga tgaagtagtt 1080
ggaaaagaca ggcttgtgca agaatcagac atacgaaatc taaactactt aaaagcttgt 1140
tgcagagaaa cattcaggat tcacccaagc gctcattatg tcccacctca tgttgcccgt 1200
caagatacca cccttggggg ttattttatt cccaaaggta gccacattca tgtatgccgc 1260
cctgggctag gccggaaccc taaaatatgg aaagatccat tagcatacga accggagcgt 1320
cacctccaag gagacggaat cacaaaagag gttactctgg tcgaaacaga gatgcgtttt 1380
gtctcattta gcactggtag acgtggctgc gtcggtgtca aagtcgggac aattatgatg 1440
gctatgatgt tggctaggtt tcttcaaggt tttaactgga aactccatcg agatttcgga 1500
ccgttaagcc tcgaggaaga tgatgcatca ttgcttatgg ctaagcctct tcttttgtct 1560
gttgagccac gcttggcatc aaacctttat ccaaaattcc gtccttaa 1608